Discovery and Characterization of the Cylindrocyclophane Biosynthetic Pathway

Discovery and Characterization of the Cylindrocyclophane Biosynthetic Pathway

A dissertation presented

by

Hitomi Nakamura

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

May 2015

Dissertation advisor: Professor Emily P. Balskus Hitomi Nakamura

Discovery and Characterization of the Cylindrocyclophane Biosynthetic Pathway

Abstract

Nature constructs structurally diverse, bioactive molecules using enzymes. Many enzymes catalyze synthetically challenging reactions under mild, physiological conditions. Consequently, they have long been a source of inspiration for developing biomimetic organic syntheses and methods. In addition, enzymes are increasingly being used as biocatalysts in industry. Therefore, the discovery of enzymes that catalyze chemically intriguing transformations can positively impact synthesis in multiple ways. With the recent advances in next-generation DNA sequencing technologies, we are now able to access enormous amount of genomic sequencing data, which encodes a treasure chest of new enzymatic chemistry. The challenge now is to devise a method to efficiently identify chemically interesting enzymes from this vast pool of information.

One possible solution to this problem is to study the biosynthetic pathways of structurally unique natural products, which are predicted to involve novel enzymatic reactions. The cylindrocyclophanes are a family of natural products that contain an unusual [7.7]paracyclophane core scaffold. Based on the previous work on the cylindrocyclophanes, their biosynthesis was predicted to involve an extraordinary C–C bond-formation. To discover the enzyme responsible for this chemistry, we studied the biosynthesis of the cylindrocyclophanes.

Chapter 2 describes the discovery and the validation of the cylindrocyclophane (*cyl*) biosynthetic gene cluster. The candidate *cyl* gene cluster was identified from the genomic sequence of the cylindrocyclophane producer. We formulated a biosynthetic hypothesis based on the *cyl* gene cluster annotation and biochemically characterized the functions of three enzymes (fatty acid activating enzymes CylA/CylB and type III PKS CylI) to connect the *cyl* gene cluster

to the production of the cylindrocyclophanes. In addition, feeding experiments using deuterium-labeled decanoic acid revealed that decanoic acid is a precursor to the cylindrocyclophanes.

Chapter 3 details our investigation of the type I PKS assembly line termination chemistry. We discovered that cylindrocyclophane biosynthesis involves a rare interaction between type I and type III PKSs, in which type III PKS CylI catalyzes assembly line termination by directly using a substrate tethered to the type I PKS CylH. Interestingly, CylH contains a C-terminal TE domain, a domain that catalyzes product release in typical assembly line enzymes. We found that the CylH TE domain has sequence and functional homology to type II editing thioesterases using both bioinformatic analyses and biochemical characterizations.

Chapter 4 details our discovery of a new halogenating enzyme CylC that catalyzes a cryptic chlorination in cylindrocyclophane biosynthesis. Bioinformatic analyses revealed that CylC resembles ferritin-like di-metallo carboxylate enzymes and that CylC and its homologs are likely responsible for chlorination of cyanobacterial natural products. CylC catalyzed chlorination of decanoyl-CylB *in vivo*, which indicated that cylindrocyclophane biosynthesis requires pre-functionalization of an unactivated carbon center for eventual C–C bond formation.

Chapter 5 discusses our discovery of a new alkylating enzyme CylK that catalyzes a Friedel–Crafts-type C–C bond formation. CylK catalyzed the final dimerization step to construct the [7.7]paracyclophane core of the cylindrocyclophanes using chlorinated resorcinol substrates, confirming that the biosynthesis proceeds through cryptic chlorination. The expanded substrate scope of CylK suggests that this enzyme is a promising candidate for future bioengineering efforts to develop useful biocatalysts for C–C bond formation. Thus, we succeeded in discovering a new halogenase and an alkylating enzyme through our investigation of cylindrocyclophane biosynthesis using a chemically guided approach.

Acknowledgements

I would like to convey my sincere gratitude to everyone who has assisted me throughout my graduate studies. I would like to thank my mentor Dr. Ellen M. Sletten and Professor Carolyn R. Bertozzi for encouraging me to pursue an undergraduate research at UC Berkeley. Experience I gained through undergraduate research greatly helped me with graduate study. I thank Dr. John Jewett for teaching me useful skills in organic synthesis during my time in the Bertozzi lab.

I would like to express my deepest gratitude to my dissertation advisor Professor Emily P. Balskus for her guidance over the last five years and for helping me become a better scientist. Emily also carried out the preliminary work on this dissertation work during her years as postdoctoral researcher in the laboratory of Professor Christopher T. Walsh. I would like to thank my Graduate Advising Committee members Professor Daniel Kahne and Professor Jon Clardy for their advice and guidance. I thank Professor Christopher T. Walsh for his support and helpful advice on this dissertation work. I would like to thank Professor Yoshito Kishi for his valuable career advice.

I also acknowledge everyone who conducted the earlier work on this dissertation research. Hilary A. Hamer performed the biochemical characterizations of CylA and CylB. Dr. Gopal Sirasani performed the synthesis of the resorcinol standard as well as some of the substrates for the feeding experiments. I thank Katarzyna Hojczyk, Ivan Bochkov and Smaranda Bodea for help with the early experiments. I acknowledge the rotation student Zebulon Levine for the synthesis of acetyl-SNAC.

I would like to convey my appreciation to everyone I had a chance to work closely with during my graduate study. I thank Dr. Pedro N. Leão for the investigation of the bartoloside biosynthesis and for the collaboration on the screen of the halogenase diversity in cyanobacteria. Towards the end of my graduate work, I had the pleasure of working with Samantha Cassel and Dr. Erica Schultz. I enjoyed working with Samantha who joined the project on the new halogenases. Erica helped me with the syntheses of some of the substrates for CylK, and she will continue the work on the mechanism and the biocatalytic

applications of CylK. I thank Jonathan Marks who is an undergraduate student I worked with on a separate project for two years. Jonathan was always enthusiastic about science and he helped me with the bioinformatics for the thioesterase work. I am grateful to Pedro and Erica for editing my thesis, and the discussions I had with them helped me greatly with my dissertation work.

I would like to thank everyone who has helped me with the techniques used for this dissertation work. I received assistance with LC-MS experiments from Professor Alan Saghatelian, Dr. Nawaporn Vinayavekhin, and Dr. Tejia Zhang for the earlier work. I thank the staff members of the Small Molecule Mass Spectrometry Facility, Dr. Sunia Trauger, Dr. Jennifer X. Wang, Dr. Gary Byrd, and Dr. Charles Vidoudez for the LC-MS analysis. I thank Dr. Adam Brown and Andrew Bendelsmith in the Jacobsen group for their help with the chiral HPLC analysis. I thank Li Zha for the help with the gel autoradiography experiments and Dr. Matthew Wilson for the help with the COSY and HSQC NMR experiments. I thank Professor Greg Challis for his advice on the synthesis of the stereselectively labeled decanoic acids.

I would like to thank my friends and colleagues who supported me during my graduate study. Smaranda Bodea is the best friend and colleague I could ever ask for and she has helped me the entire time we spent together in the Balskus group. I thank my friends in the department Fan Liu and Kimberli Kamer who motivated me with their enthusiasm for science. I am grateful to my friend Carissa Pardamean for keeping me company throughout my years in undergraduate and graduate studies.

Finally, I would like to express my gratitude to my family who has always supported me and encouraged me. I am thankful to my sister Dr. Yuki Nakamura who has been my role model since I was little. She is also a talented chemist and I had the pleasure of living with her during my second and third years in graduate study while she worked as a postdoctoral researcher in the Kishi group. I thank my wonderful parents Kaoru and Noriko Nakamura for helping me in all possible ways. Their unwavering love, support and guidance continue to be a driving force in my life.

Table of Contents

List of Abbreviations

ºC	degree Celsius
Ci	curie
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
COSY	correlation spectroscopy
d	deuterium
DDM	*n*-dodecyl β-D-maltoside
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDC•HCl	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide•hydrochloride
eV	electron volt
EPPS	4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid
g	gram
h	hour
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	coupling constant
L	liter
LC-MS	liquid chromatography-mass spectrometry
M	molar
min	minute
mol	mole

MS/MS	tandem mass spectrometry
MWCO	molecular weight cut-off
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
OD	optical density
psi	pounds per square inch
PCC	Pasteur Culture Collection of Cyanobacteria
PCR	polymerase chain reaction
rpm	revolutions per minute
s	second
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
UV	ultraviolet
V	volt
Vis	visible
□g	times gravity

Chapter 1. Introduction to the discovery of biosynthetic pathways and the cylindrocyclophanes

1.1. Introduction to the discovery of biosynthetic pathways and genome mining

Nature constructs a wide variety of complex organic molecules from simple building blocks. The classes of natural products that have been isolated thus far have staggeringly diverse chemical scaffolds; some of the major families are the polyketides, nonribosomal peptides, saccharides, alkaloids and terpenoids (**Figure 1.1**).[1] Scientists have long been interested in questions surrounding these complex bioactive molecules, such as how and why are they made. The study of the biological activities of natural products has been intertwined with their potential therapeutic benefits; over the past 30 years, 61% of anticancer drugs and 49% of antimicrobial agents are either natural products or compounds derived from them.[2]

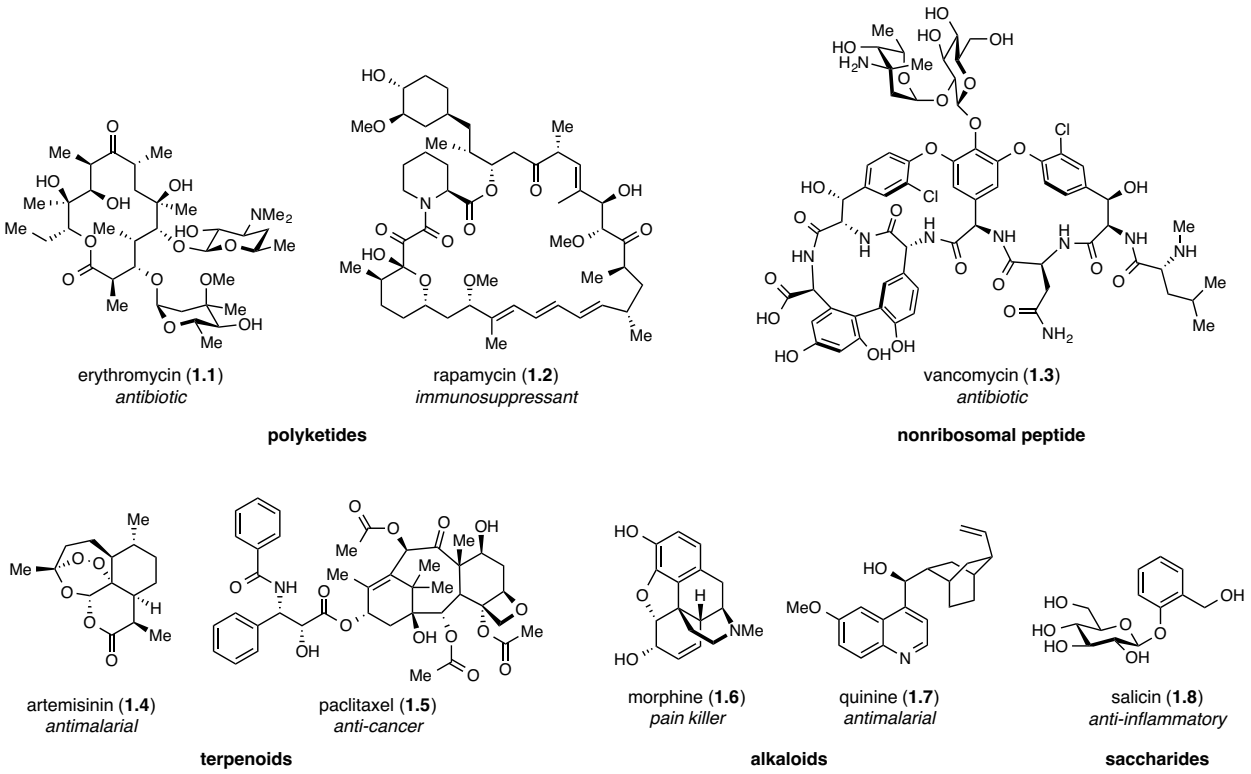


Figure 1.1: Natural products with pharmacological properties.

Microorganisms are prolific producers of bioactive molecules. Intriguingly, most microbially derived natural products are assembled by enzymatic pathways that are encoded by group of adjacent genes

known as the biosynthetic gene clusters.[3] This physical clustering of the natural product biosynthetic genes is a significant advantage to studying biosynthetic pathways in microorganisms. Recent evolution of the DNA-sequencing technologies has facilitated rapid, low-cost microbial genome sequencing, resulting in a steep increase in the amount of available genomic data. In fact, the number of sequenced bacterial genomes has exponentially increased since the first bacterial genome was sequenced in 1995. By 2015, more than 30,000 sequenced bacterial genomes were deposited in public databases, and the number continues to rise rapidly.[4] In addition, recent progress in metagenomics[5] and single-cell genome sequencing[6] provides access to genetic information even of microbial strains that are not readily culturable.

Growth in bioinformatics has accompanied this expansion of accessible genomic data: more and more computational tools are being developed for analysis of DNA-sequencing data and gene identification.[3] The process of using computational tools to extract out valuable information on gene functions and biosynthetic gene clusters from DNA-sequencing data is termed "genome mining".[7] In addition to the simple gene and protein comparison tools such as Basic Local Alignment Search Tool (BLAST),[8] increasing numbers of search tools for identification and analysis of biosynthetic gene clusters such as antiSMASH,[9,10] NaPDoS[11] and NP.searcher[12] are becoming publicly available. While these online tools are capable of providing a good starting point for biosynthetic cluster identifications and annotations, in many cases they are not yet sophisticated enough to provide good predictions for the structures of the metabolites produced by the cluster.

1.1.1. Connecting sequencing data to natural product biosynthesis

With the massive influx of genetic information, efforts to connect the DNA-sequencing data to the actual natural products being produced are also progressing. Biosynthetic gene clusters that have not been connected to the metabolites they produce are called "orphan clusters", and methods such as the genomisotopic approach are being developed to identify the natural products made by orphan clusters.[13] Another method of connecting natural products to orphan clusters that is gaining popularity is the

metabolomics approach that matches genes to molecules using mass spectrometry. Algorithms such as RiPPquest[14] and NRPquest[15] allow computational coupling of mass spectra and genomic data for peptidic natural products, such as nonribosomal peptides and ribosomally synthesized and post-translationally modified peptides (RiPPs). This approach, however, is not applicable to other classes of natural products for which the mass spectrometry fragmentation patterns are more challenging to predict.

In addition to the tools available for predicting metabolites produced by biosynthetic gene clusters, there are also bioinformatic tools for connecting gene sequences to functionally characterized proteins and protein families. One complementary tool to BLAST is HHpred, which predicts secondary structures from primary protein sequences and creates structural alignments with a set of templates selected from databases such as the Protein Data Bank (PDB).[16] Comparison of secondary structure allows identification of proteins that belong to protein families with conserved structural features but distant amino acid sequences. Another computational tool that is gaining popularity is the sequence similarity networks that allows visualization of functional trends across protein superfamilies based on sequence similarity.[17] Use of these computational tools can facilitate future efforts to predict protein structures and functions from sequencing data.

As the number of sequenced genomes as well as metagenomic sequences continues to rise, the next major challenge is to devise a method to efficiently extract out useful information from these data. Although numerous bioinformatic methods are being developed for better organization of sequencing data and more accurate analyses, there is still a long way to go until we can fully rely on computational tools to connect DNA-sequences to natural product structures or predict enzyme functions. A critical bottleneck in this process is the experimental characterization step to verify whether the computational analyses are correct and to assign functions to currently unknown genes and proteins. Because of the substantial time and effort experimental characterization step requires, being able to rapidly identify the biosynthetic pathways of greatest biological or chemical interest is essential.

1.1.2. Using chemical knowledge to discover novel microbial enzymes

One important motivation for studying the biosynthesis of natural products is to discover the functions of novel enzymes. Enzymes are nature's tools for synthesis of organic molecules that have evolved over billions of years to construct stunningly complex natural products. Enzymes have long been a source of inspiration for chemists, leading to development of biomimetic methodologies and total syntheses.[18-20] Many enzymes are capable of catalyzing synthetically challenging transformations with superb selectivity and specificity under mild physiological conditions. In recent years, biocatalytic approaches are increasingly employed in drug and fine-chemical industries as more environmentally friendly and economical alternatives to traditional synthetic methods.[21]

The revolutionary advances in DNA-sequencing technologies and computational tools provide chemists with the opportunity to uncover synthetically useful biocatalysts from a vast pool of enzyme sequences. The major challenge in reaching this goal is to devise a means to efficiently identify enzymes that catalyze new and desirable transformations. While bioinformatic tools excel at predicting functions of gene products with homology to previously characterized enzymes, many novel enzymes remain misannotated or labeled as "hypothetical proteins".[22] This reliance on previous knowledge poses a challenge for discovering new chemistry catalyzed by truly novel enzymes lacking close homologs. In fact, it has been indicated that less than 1% of the annotated microbial genes have been experimentally characterized.[23] As a consequence, our ability to discover new enzymatic chemistry based solely on DNA-sequencing data is limited.

One possible solution to overcoming this hurdle is to apply knowledge of organic chemistry to identify biosynthetic pathways with intriguing chemical transformations and link sequencing data with enzyme functions. For instance, chemists can generate biosynthetic hypotheses based on retrosynthetic analyses of the target molecules, which can help guide the search for biosynthetic gene clusters from microbial genome sequencing data. Moreover, chemical knowledge can also be applied to experimental designs for assigning important biological functions to previously unrecognized genes and proteins. In chemically-

guided discovery of biosynthetic gene clusters, chemists can recognize intriguing structures and apply reactivity principles that direct their bioinformatic searches. Once the candidate gene cluster is identified, pathway logic and enzyme functions can be applied to formulate a biochemical hypothesis (**Figure 1.2**). In this work, we have employed chemically guided approaches for the discovery and the characterizations of the biosynthetic pathway of a family of dimeric polyketides containing a unique paracyclophane core.

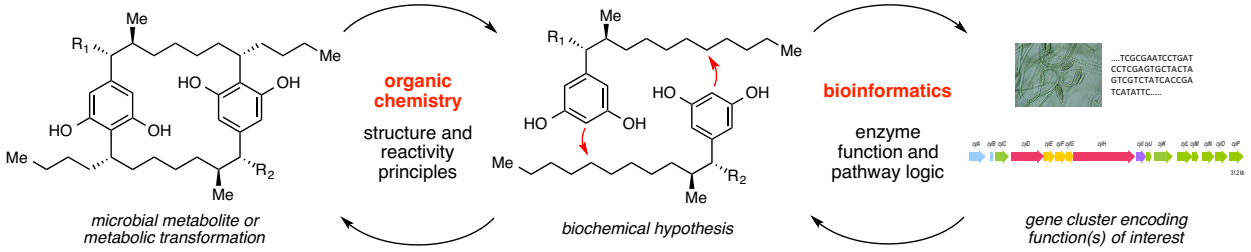


Figure 1.2: Chemically guided approaches for discovering and understanding biological function in microbial genome sequencing data.

1.2. Introduction to the biosynthesis of polyketides

Polyketides are a class of natural products with remarkable structural diversity and biological activities.[24] Many polyketides such as erythromycin (**1.1**), rapamycin (**1.2**) and tetracycline exhibit pharmacologically important properties, including antibiotic, anticancer, antifungal, antiparasitic and immunosuppressive activities. As the name indicates, polyketides are derived from polyketones that are generated from acetate units through repeated condensation reactions, which was first confirmed in the 1950s by the incorporation of the radiolabeled acetate units into 6-methylsalicylic acid produced by the fungus *Penicillium patulum*.[25] Since then, the field of biosynthesis expanded rapidly with the development of nuclear magnetic resonance (NMR) and mass spectroscopy (MS), and scientists have used building blocks labeled with stable isotopes (namely ^{13}C, ^{2}H, ^{18}O and ^{15}N) to investigate the origins of polyketides.[26]

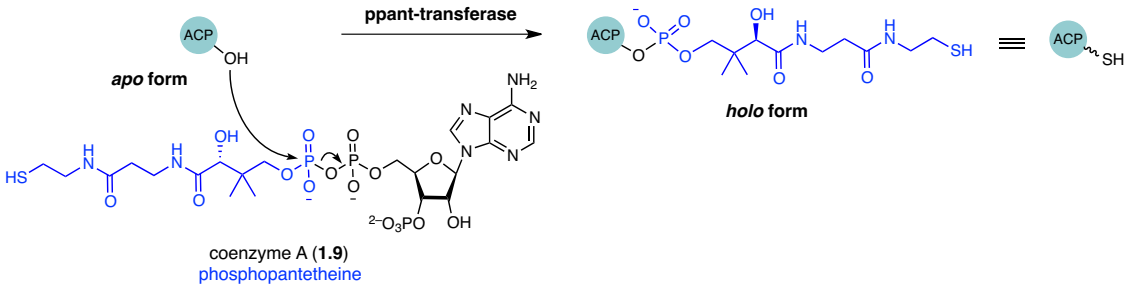
Polyketides are generated by enzymes called polyketide synthases (PKSs). While the biosynthetic origins of polyketides were actively studied in the 1960s and the 1970s, the actual enzymes that assemble these

molecules remained elusive for decades. The efforts to characterize the PKSs dramatically accelerated with the development of genetic techniques in the 1980s, and the first pioneering genetics work was performed by Malpartida and Hopwood on the polyketide actinorhodin biosynthesis.[27] Several years later, the Leadlay group and the Katz lab independently discovered the PKS enzymes responsible for the production of erythromycin.[28-31] The first genetic studies on the actinorhodin and the erythromicn biosyntheses revealed that the PKSs have high functional homology to fatty acid synthases.[32] The research on polyketide biosynthesis has vastly expanded since the initial studies in the late 1980s and the early 1990s, and the functions and diversity of the PKS enzymes have been investigated in detail.

1.2.1. Functional domains of polyketide synthases

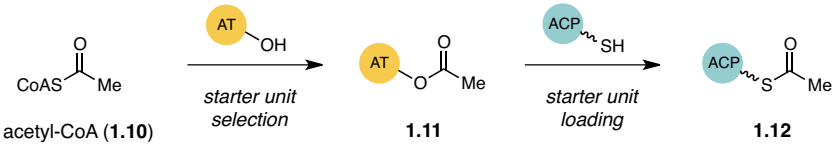
PKS enzymes consist of functional units or domains, each with distinct roles in polyketide assembly. The core domains of PKS enzymes are the ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) domains.[33]

The function of the ACP domain is to carry the covalently tethered substrate from one domain to another as the molecule is being modified. Prior to catalysis, the ACP domain must be post-translationally modified from the inactive *apo* form to the active *holo* form. The activation of the *apo* protein to the *holo* protein is catalyzed by a ppant-transferase, which attaches a phosphopantetheinyl (ppant) arm derived from coenzyme A (**1.9**) to a conserved serine residue on the ACP (**Scheme 1.1**).[34] A biosynthetic gene cluster of secondary metabolite may encode a dedicated ppant-transferase, but in some cases the ppant-transferase from primary metabolism or other biosynthetic gene cluster catalyzes this post-translational modification.[35,36] Due to this covalent tethering of the substrate during structural modifications, the PKSs and non-ribosomal peptide synthetases (NRPSs) are often considered as “assembly line enzymes”.[37]

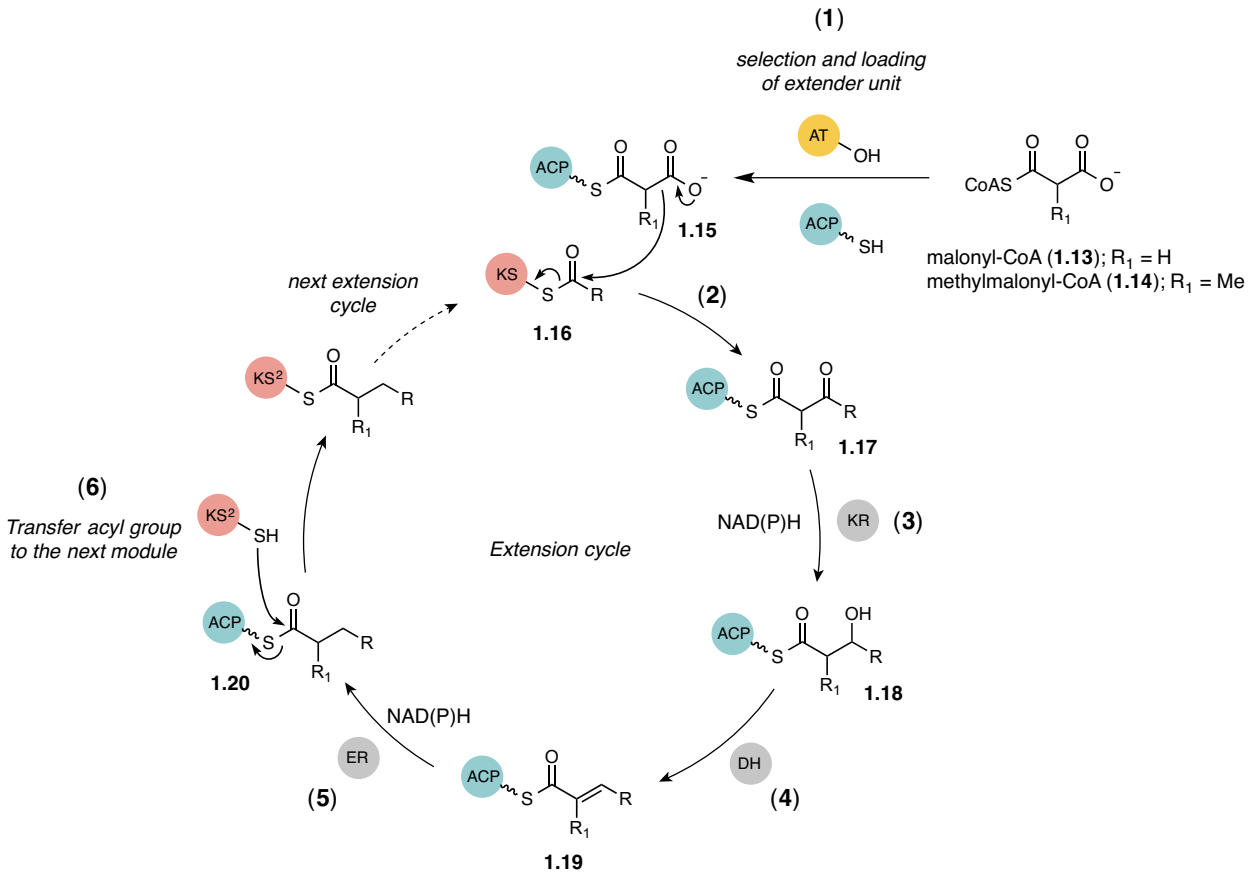


Scheme 1.1: Post-translational modification of the ACP catalyzed by the ppant-transferase

The AT domains are responsible for the selection and the activation of starter units or extender units during the initiation step or the extension cycle in PKS assembly lines. Typically, AT domains that select starter unit activates acyl-CoA thioesters such as acetyl-CoA (**1.10**) in the initiation step (**Scheme 1.2**). The activation of the monomer unit is catalyzed by a serine residue of the AT domain to form a covalent acyl-AT (**1.11**) intermediate. Once the AT domain selects and activates a monomer, it then loads the activated monomer onto the thiol of the ppant arm of the ACP domain (**1.12**). The intermediate transferred to the ACP domain by the AT domain is covalently linked to the ppant arm by a thioester linkage. The AT domains that select and activate extender units during extension cycles (**Reaction 1**, **Scheme 1.3**) typically use malonyl-CoA (**1.13**) or methylmalonyl-CoA thioesters (**1.14**). For the AT domains that activate malonyl-CoA (**1.13**) or methylmalonyl-CoA (**1.14**) extender units, the monomer specificity can often be predicted from the amino acid sequence of the domain.[38]



Scheme 1.2: A representative PKS assembly line initiation step catalyzed by the AT domain.



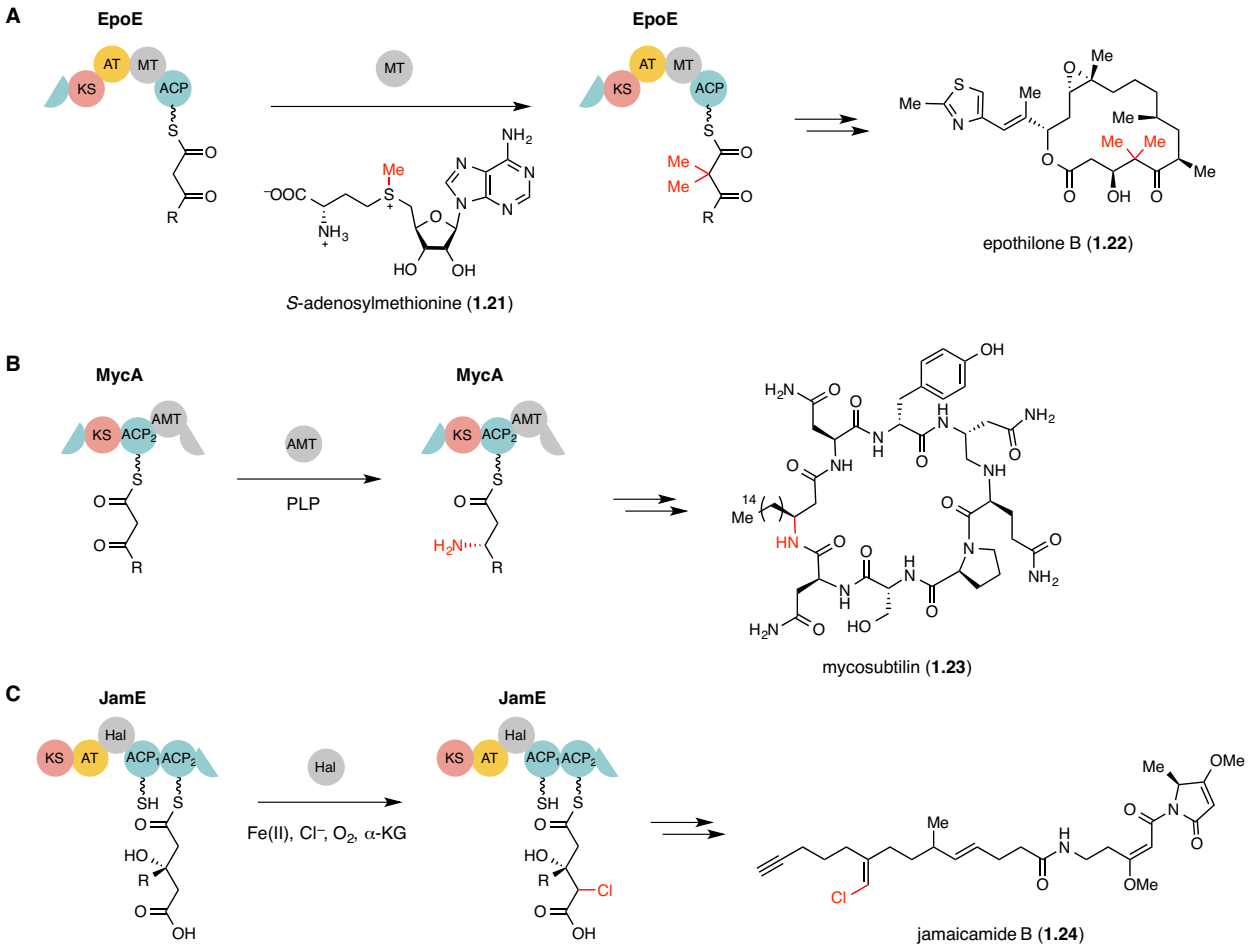
Scheme 1.3: A representative PKS extension cycle showing the reactions catalyzed by the AT, KS and ACP domains as well as the KR, DH and ER tailoring domains.

After the loading of the ACP domain, the KS domain catalyzes the actual C–C bond formation through a decarboxylative Claisen condensation reaction (**Reaction 2, Scheme 1.3**). The KS domain first receives the substrate from the upstream ACP domain by forming a thioester intermediate with its catalytic cysteine residue (**1.16**), and the thioester then serves as an electrophile for the nucleophilic attack by the downstream ACP-tethered malonate unit (**1.15**). The timed decarboxylation of the malonate extender unit provides thermodynamic driving force and renders the C–C bond forming steps irreversible.[33]

The Claisen-condesation catalyzed by the KS domain results in a □-ketothioester product (**1.17**), which can be further modified by tailoring domains that alter the functionality and oxidation state at the □-position. These additional tailoring domains are the ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains. The KR, DH and ER domains work sequentially to transform a □-ketone moiety

into a methylene, through reduction of the ketone to an alcohol (**1.18**) by the KR domain using NAD(P)H (**Reaction 3**, **Scheme 1.3**), dehydration by the DH domain to form a *trans* alkene (**1.19**, **Reaction 4**, **Scheme 1.3**), and reduction of the alkene to a methylene (**1.20**) by the ER domain using NAD(P)H (**Reaction 5**, **Scheme 1.3**). The oxidation state of the product may vary if only the KR or KR and DH domains are present. The alteration of the oxidation state by the additional tailoring domains contributes to the structural diversity of polyketides.

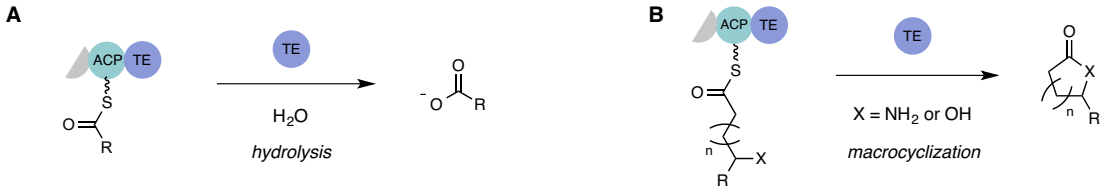
Besides the domains mentioned above, some PKS enzymes also contain less common domains that provide even more chemical diversity to the polyketide structure.[39] For instance, some PKS enzymes also contain a methyltransferase (MT) domain, which installs an α-methyl substituent using *S*-adenosylmethionine (SAM, **1.21**) as a methyl donor as in the case of the epothilone (**1.22**) biosynthesis (**Scheme 1.4A**).[39] Other examples of unusual tailoring domains include the aminotransferase (AMT) domain that uses a pyridoxal 5'-phosphate (PLP) cofactor to catalyze reductive amination of the β-ketone moiety in the mycosubtilin (**1.23**) biosynthesis (**Scheme 1.4B**),[40] and the halogenase (Hal) domain that catalyzes chlorination in the curacin and the jamaicamide (**1.24**) biosyntheses (**Scheme 1.4C**).[41]



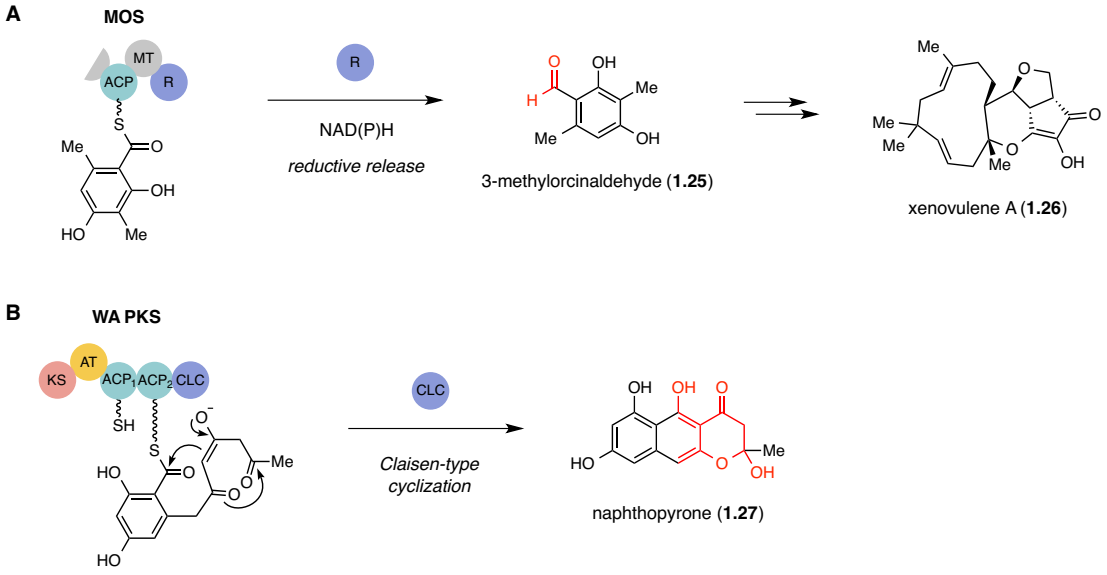
Scheme 1.4: Examples of unusual tailoring domains found in PKS assembly lines. **A**) *C*-methylation by a methyltransferase domain. **B**) Reductive amination by an aminotransferase domain. **C**) Chlorination by a halogenase domain.

Another important domain in the PKS systems is the thioesterase (TE) domain, which is responsible for releasing the nascent polyketide at the end of the assembly line. The TE domain can use water as a nucleophile to hydrolyze the acyl-ACP thioester linkage to form a free acid (**Scheme 1.5A**) or use a nucleophilic alcohol or amine group on the substrate to form a macrolactone or a macrolactam (**Scheme 1.5B**).[42] In some cases, PKS enzymes carry other domains in place of the TE domain for product release. For example, reductase (R) domains have been shown to catalyze NAD(P)H-dependent reductive release in pathways such as the xenovulene (**1.26**) biosynthesis (**Scheme 1.6A**).[43] The R domain can catalyze either a two-electron reduction to form an aldehyde or a four-electron reduction to form an alcohol. Another commonly found release domain in the fungal PKS pathways is the Claisen-like cyclase (CLC)

domain. The CLC domain uses a carbon nucleophile to attack the carbonyl of the acyl-ACP thioester to form a new aromatic ring system (**Scheme 1.6B**). Examples of the product release catalyzed by the CLC domain are found in the biosyntheses of polyketides such as naphthopyrone (**1.27**)[44] and aflatoxin.[45] The termination process is an important step in the PKS assembly line that contributes to the chemical diversity of the polyketides.



Scheme 1.5: Reactions by typical C-terminal TE domains. **A**) Hydrolysis reaction. **B**) Macrocyclization to form macrolactam or macrolactone.

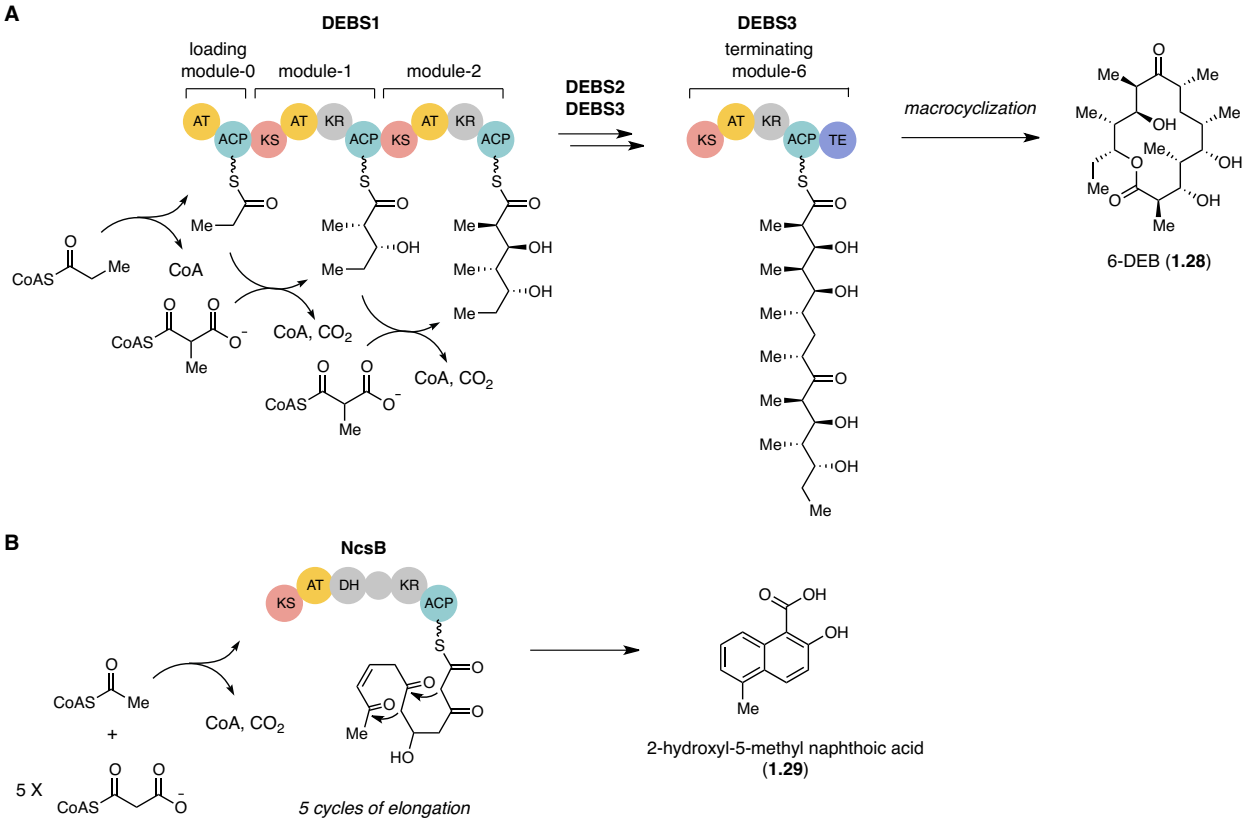


Scheme 1.6: Other commonly observed assembly line termination. **A**) Reductive release by reductase domain. **B**) Product release through cyclization catalyzed by Claisen-like cyclase domain.

1.2.2. Types of polyketide synthases

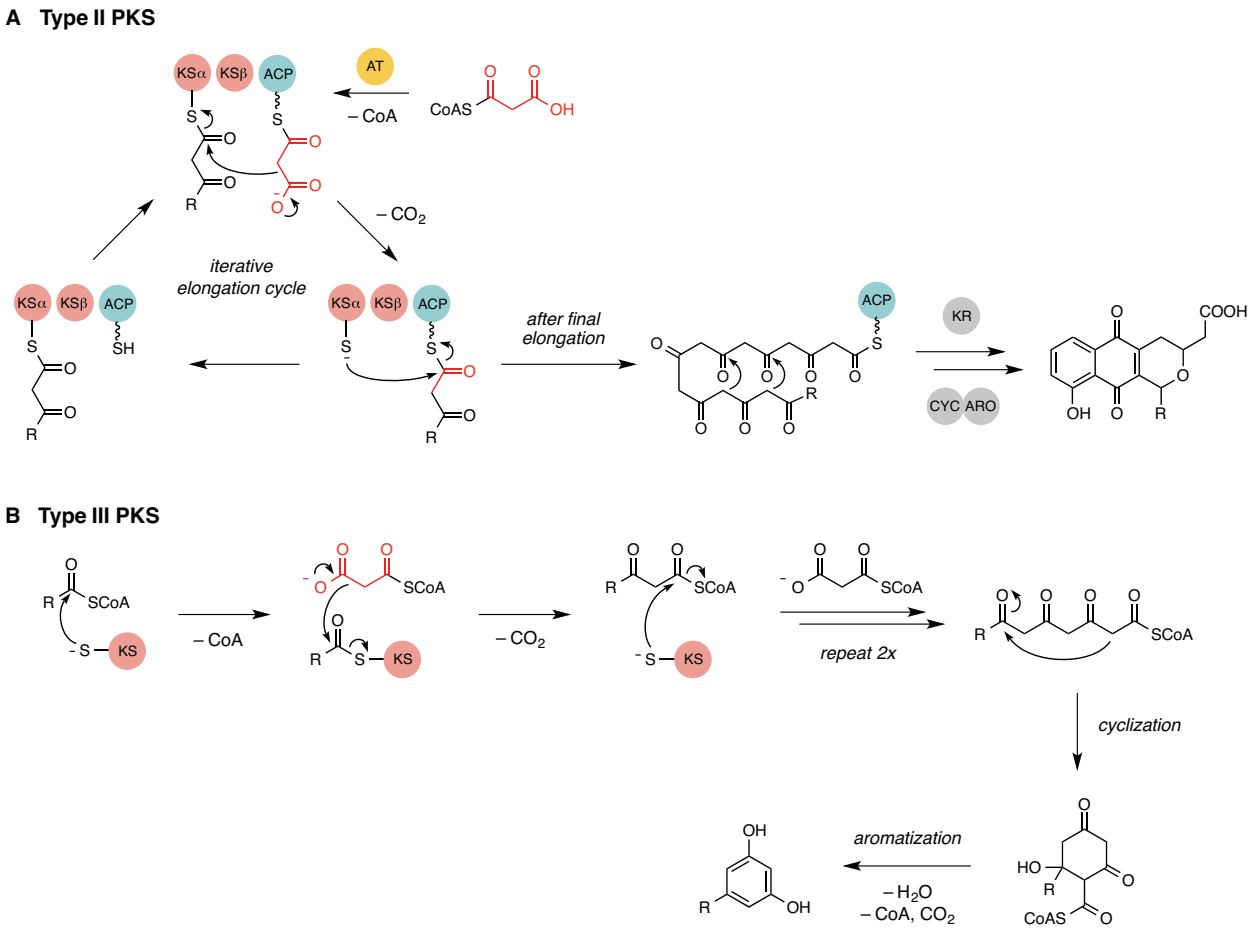
PKS enzymes are classified into different types based on how the domains are organized (**Scheme 1.7**).[46] Type I PKSs are large multifunctional enzymes that contain one or more modules, each of which consist of multiple domains as a part of large polypeptides. A typical module of type I PKSs carries KS, AT and

ACP domains as well as variations of the tailoring KR, DH and ER domains. The loading module, or the initial starting module responsible for the activation and loading of an acyl-CoA substrate onto the type I PKS assembly line, lacks a KS domain and typically only consists of AT and ACP di-domains. Furthermore, the type I PKSs can be classified as either modular or iterative. Each module of the type I modular PKSs catalyzes only one cycle of polyketide chain elongation, as in the case of the 6-deoxyerythromycin B synthase (DEBS) (**Scheme 1.7A**).[47] On the other hand, a single module of the type I iterative PKSs can catalyze multiple rounds of elongation. In some cases, not every domain is used in each elongation cycle, as in the case of 2-hydroxyl-5-methyl naphthoic acid (**1.29**) formation by NcsB (**Scheme 1.7B**).[48] The type I modular PKSs are found in bacteria, while type I iterative PKSs are more commonly found in fungal biosynthetic pathways.[49] Type I iterative PKSs can be further subdivided into non-reducing, partially reducing, and highly reducing PKSs.[50]



Scheme 1.7: Examples of the type I PKSs. **A**) Type I modular PKSs in DEBS biosynthesis. **B**) Type I iterative PKS in naphthoic acid biosynthesis.

Unlike the type I PKSs with physically linked domains, the type II PKSs are multienzyme complexes. In other words, each domain is a freestanding enzyme in the type II PKSs, and these enzymes usually function iteratively (**Scheme 1.8A**). The type II PKSs produce polyphenolic ring systems such as anthracyclines, angucyclines and tetracyclines, and Gram-positive actinomycetes are a particularly rich source of these polyketides.[51] A minimal type II PKS consists of two KS units (KS-α and KS-β) and an ACP, which are typically found adjacently in a gene cluster. The KS-α and KS-β subunits form a heterodimer: the KS-α subunit catalyzes decarboxylative Claisen condensations while the KS-β subunit lacks a catalytic cysteine residue and is thought to control the chain length or number of iteration.[51] Due to its function, the KS-β subunit is also called the chain length factor (CLF).[52] Additional domains of the type II PKSs include KR, cyclase (CYC), and aromatase (ARO). The KR domain tunes the oxidation state of the polyketone intermediate to form alcohols. The CYC domain is a chaperone-like protein that controls the initial ring formation through aldol reactions, and the ARO domain catalyzes the aromatic ring formation through series of dehydration reactions. CYC and ARO have been found either as a fused di-domain protein or as separate monodomain protein. Together these additional domains define the folding pattern of the nascent polyketone intermediate to form aromatic ring structures that are characteristic of type II polyketides.[51]



Scheme 1.8: Representative catalytic cycles of type II and type III PKSs. **A**) General reactions catalyzed by type II PKSs. **B**) General reactions catalyzed by type III PKSs.

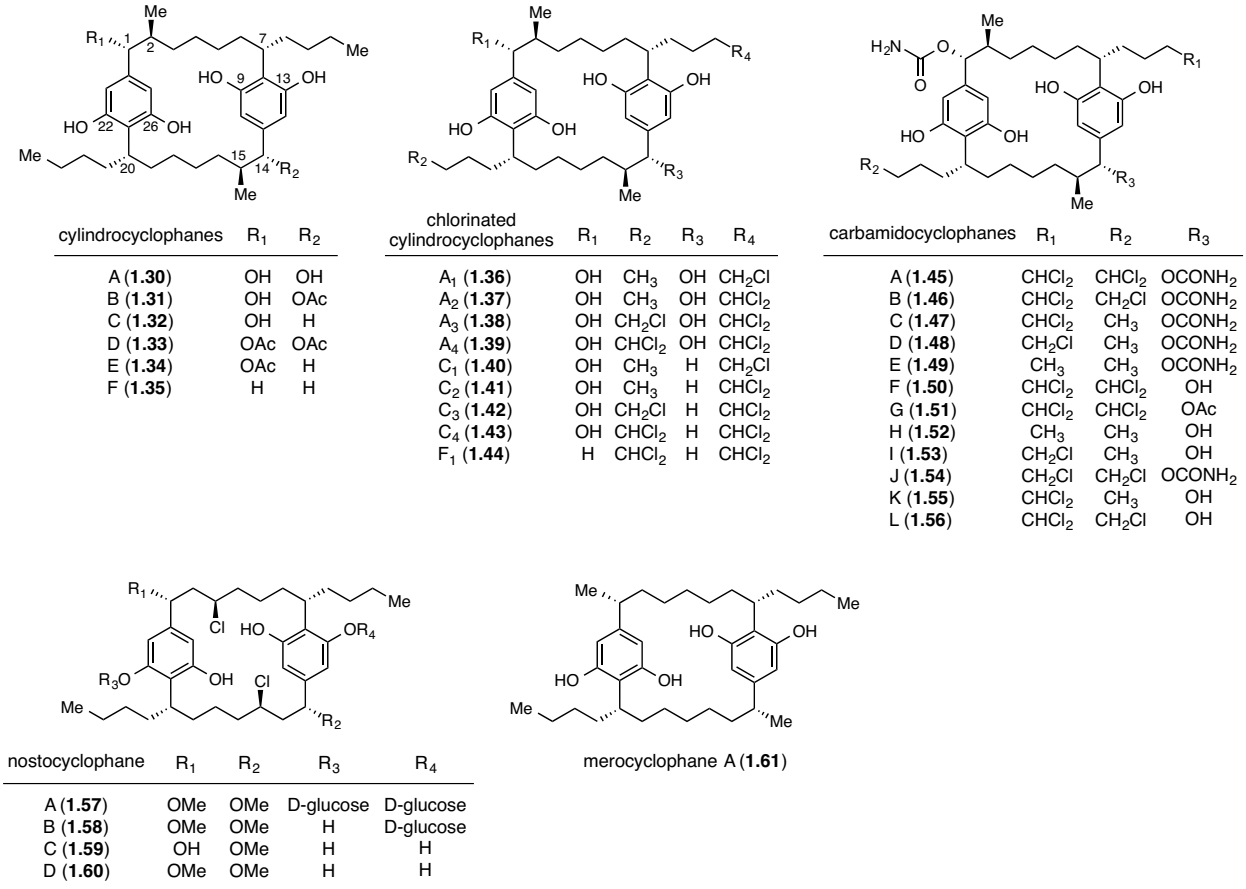
The type III PKSs are considerably simple compared to the type I and the type II PKSs. A type III PKS consists of a homodimer of identical KS monomeric domains and lacks an ACP domain for covalent tethering of the building blocks (**Scheme 1.8B**).[53] The type III PKSs typically use acyl-CoA electrophiles as substrates and catalyze iterative decarboxylative Claisen condensations directly with malonyl-CoA extender units. More recently several type III PKSs have been shown to also accept acyl-ACP substrates from fatty acid biosynthesis to produce phenolic lipids or alkylpyrones.[54,55] Although the type III PKSs only consist of a single dimeric KS domain, their catalytic functions are rather complex: these enzymes control iterative C–C bond formations, cyclizations, and dehydration steps to form aromatic polyketides.[53]

One of the most well known polyketides synthesized by a type III PKS is resveratrol, which is found in plants such as grapes and blueberries.[56]

As more polyketide biosynthetic pathways are characterized, unusual PKS enzymes that digress from previously known types of PKSs are being discovered. For instance, multiple type I PKSs that lack *in cis* AT domains and instead have a freestanding AT that acts *in trans* have been found.[57] Such PKSs are called *trans*-AT or AT-less PKSs. There have also been reports of type II PKSs that act non-iteratively, lack ACP and utilize acyl-CoAs directly as substrates.[58,59] Moreover, many biosynthetic gene clusters that contain PKSs also encode NRPSs, which are known as hybrid NRPS-PKS clusters.[60] Additional knowledge of PKS pathways and discoveries of new types of PKSs may lead to modifications of the currently used PKS classification in the future.

1.3. General introduction to the cylindrocyclophanes

Cylindrocyclophanes are a family of polyketides with a unique core structure that have attracted the attention of both chemists and biologists. The cylindrocyclophanes (**1.30**-**1.35**, see **Figure 1.3**) were first isolated in the early 1990s from the filamentous cyanobacterium *Cylindrospermum licheniforme* ATCC 29204 and these molecules were the first naturally occurring paracyclophanes to be discovered.[61,62] Since the initial discovery, numerous analogs have been isolated from related filamentous cyanobacterial strains (**Figure 1.3**). Chlorinated cylindrocyclophanes (**1.36**-**1.44**) have been isolated from *Nostoc* sp. UIC 10022A,[63] and carbamidocyclophanes (**1.45**-**1.56**) containing carbamoyl moiety have been isolated from *Nostoc* sp. CAVN10,[64] CAVN2[65] and UIC 10274.[66] Nostocyclophanes (**1.57**-**1.60**) produced by in *Nostoc linckia* UTEX B1932[67] and merocyclophane A (**1.61**) found in *Nostoc* sp. UIC 10062[68] also contain the same [7.7]paracyclophane scaffold as the cylindrocyclophanes, but with different methylation patterns.



cylindrocyclophanes	R_1	R_2
A (**1.30**)	OH	OH
B (**1.31**)	OH	OAc
C (**1.32**)	OH	H
D (**1.33**)	OAc	OAc
E (**1.34**)	OAc	H
F (**1.35**)	H	H

chlorinated cylindrocyclophanes	R_1	R_2	R_3	R_4
A_1 (**1.36**)	OH	CH_3	OH	CH_2Cl
A_2 (**1.37**)	OH	CH_3	OH	$CHCl_2$
A_3 (**1.38**)	OH	CH_2Cl	OH	$CHCl_2$
A_4 (**1.39**)	OH	$CHCl_2$	OH	$CHCl_2$
C_1 (**1.40**)	OH	CH_3	H	CH_2Cl
C_2 (**1.41**)	OH	CH_3	H	$CHCl_2$
C_3 (**1.42**)	OH	CH_2Cl	H	$CHCl_2$
C_4 (**1.43**)	OH	$CHCl_2$	H	$CHCl_2$
F_1 (**1.44**)	H	$CHCl_2$	H	$CHCl_2$

carbamidocyclophanes	R_1	R_2	R_3
A (**1.45**)	$CHCl_2$	$CHCl_2$	$OCONH_2$
B (**1.46**)	$CHCl_2$	CH_2Cl	$OCONH_2$
C (**1.47**)	$CHCl_2$	CH_3	$OCONH_2$
D (**1.48**)	CH_2Cl	CH_3	$OCONH_2$
E (**1.49**)	CH_3	CH_3	$OCONH_2$
F (**1.50**)	$CHCl_2$	$CHCl_2$	OH
G (**1.51**)	$CHCl_2$	$CHCl_2$	OAc
H (**1.52**)	CH_3	CH_3	OH
I (**1.53**)	CH_2Cl	CH_3	OH
J (**1.54**)	CH_2Cl	CH_2Cl	$OCONH_2$
K (**1.55**)	$CHCl_2$	CH_3	OH
L (**1.56**)	$CHCl_2$	CH_2Cl	OH

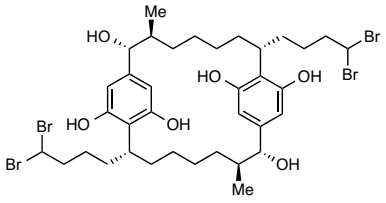
nostocyclophane	R_1	R_2	R_3	R_4
A (**1.57**)	OMe	OMe	D-glucose	D-glucose
B (**1.58**)	OMe	OMe	H	D-glucose
C (**1.59**)	OH	OMe	H	H
D (**1.60**)	OMe	OMe	H	H

merocyclophane A (**1.61**)

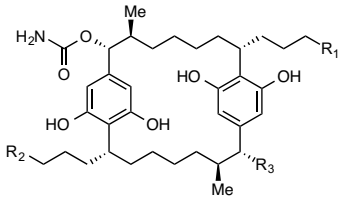
Figure 1.3: The structures of the cylindrocyclophanes and other related [7.7]paracyclophanes.

Previous work on halogenated natural products has indicated that many halogenating enzymes are relatively promiscuous in terms of the halide they can use.[69,70] To investigate the promiscuity of the cylindrocyclophane biosynthetic pathway toward different halides, Chlipala and co-workers cultured the organism that produces the chlorinated cylindrocyclophanes (**1.36**-**1.44**), *Nostoc* sp. UIC 10022A, in a chloride-deficient and bromine-enriched media. Addition of KBr to chloride-deficient media resulted in the production of brominated cylindrocyclophanes A_{B4} (**1.62**, see **Figure 1.4**) containing bromine atoms in place of the chlorine atoms. More recently, Preisitsch and co-workers also repeated the same halide replacement culturing with the carbamidocyclophane-producing strain, *Nostoc* sp. CAVN2.[71] Consistent with the results obtained by Chlipala and co-workers, the production of brominated carbamidocyclophanes M-U (**1.63**-**1.71**) was observed in bromide-enriched, chloride-deficient media.

Iodinated and fluorinated analogs were not produced when the culture media was supplemented with KI and KF, indicating that the halogenating enzyme(s) in the pathway can only accept chloride and bromide as substrates.



cylindrocyclophane A_{B4} (**1.62**)



carbamidocyclophanes	R_1	R_2	R_3
M (**1.63**)	CH_2Br	CH_3	$OCONH_2$
N (**1.64**)	CH_2Br	CH_2Br	$OCONH_2$
O (**1.65**)	$CHBr_2$	CH_3	$OCONH_2$
P (**1.66**)	$CHBr_2$	CH_2Br	$OCONH_2$
Q (**1.67**)	$CHBr_2$	$CHBr_2$	$OCONH_2$
R (**1.68**)	CH_2Br	CH_3	OH
S (**1.69**)	$CHBr_2$	CH_3	OH
T (**1.70**)	$CHBr_2$	CH_2Br	OH
U (**1.71**)	$CHBr_2$	$CHBr_2$	OH

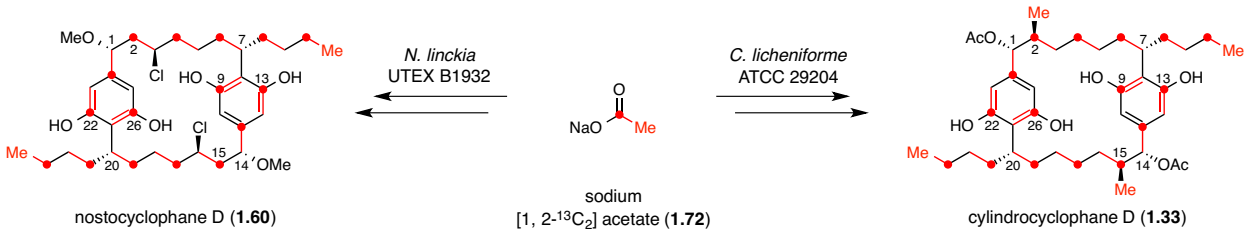
Figure 1.4: The structures of the brominated cylindrocyclophanes and carbamidocyclophanes produced under bromine-enriched conditions.

The cylindrocyclophanes (**1.30**-**1.44**) and their analogs (**1.45**-**1.61**) display a wide range of bioactivities including cytotoxicity and antiproliferative activity against cancer cell lines,[62-68] antimicrobial activities against Gram-positive bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *Mycobacterium tuberculosis*.[65,66,71] Furthermore, chlorinated cylindrocyclophanes A_2-A_4 (**1.37**-**1.39**) and A_{B4} (**1.62**) have notable 20S proteasome-inhibitory effects, which indicates that these molecules may be useful for treatment of cancers.[63] With diverse bioactivities and unique structural features, cylindrocyclophanes and related compounds are attractive target molecules for biosynthetic studies.

1.3.1. Previous feeding studies and the original biosynthetic hypothesis

Soon after the discovery of the cylindrocyclophanes (**1.30**-**1.35**) and the nostocyclophanes (**1.57**-**1.60**), Bobzin and Moore conducted feeding experiments with isotopically labeled sodium acetate and L-methionine in the native producer, *Cylindrospermum licheniforme* ATCC 29204 (also known as *Cylindrospermum stagnale* PCC 7417) and *Nostoc linckia* UTEX B1932, respectively. Feeding of sodium [1,2-$^{13}C_2$] acetate (**1.72**) resulted in the incorporation into nostocyclophane D (**1.60**) and

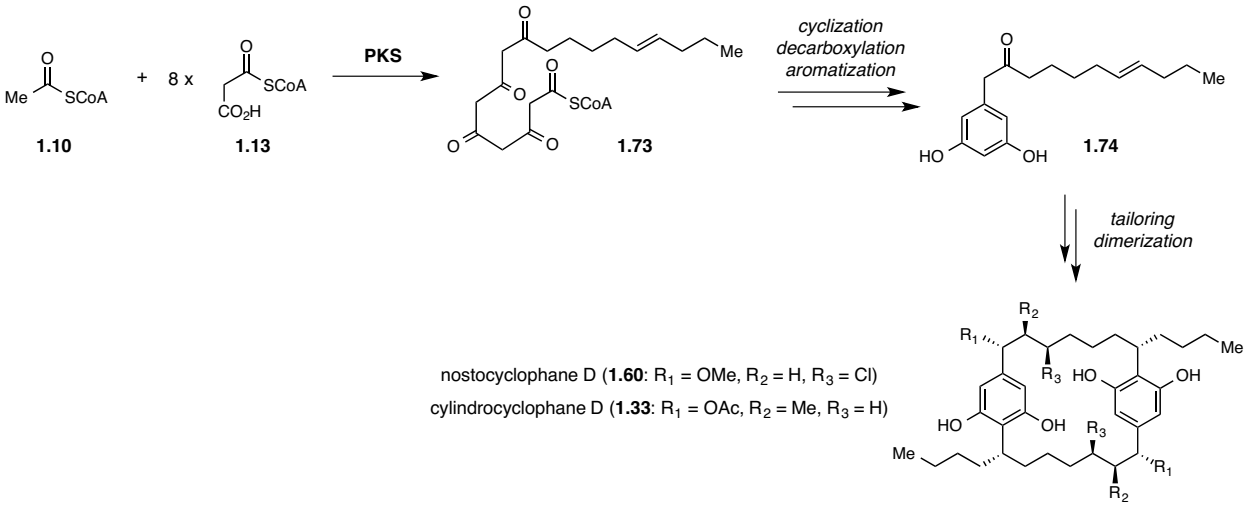
cylindrocyclophane D (**1.33**), and the labeling pattern observed suggested that these compounds are dimeric molecules consisting of two identical monomers (**Scheme 1.9**). The feeding experiment was also performed with other isotopically labeled acetates (sodium[1-^{13}C,^{2}H$_3$]acetate, sodium[2-^{13}C,^{2}H$_3$]acetate, sodium[^{2}H$_3$]acetate, sodium[1-^{13}C,^{18}O$_2$]acetate and sodium[2-^{13}C,^{18}O$_2$]acetate) to confirm the observed labeling pattern.[72] Each subunit appears to be biosynthesized in a typical polyketide fashion whereby eight malonate units are condensed sequentially with a starter acetate unit to form a nonaketide. Carbons from the last four acetate units of the nonaketide are cyclized to form the aromatic resorcinol ring and the last carbon atom in the chain is apparently lost by decarboxylation.[72] No significant ^{13}C–^{13}C coupling was observed between C7/C20 and C8/C21, indicating that this is the likely disconnection site for the dimerization.



Scheme 1.9: Sodium [1,2-^{13}C$_2$]acetate feeding study in the native producer indicates that the cylindrocyclophanes and the nostocyclophanes are dimeric polyketides.

Other feeding experiments using sodium [2-^{13}C,^{18}O$_2$]acetate showed that the oxygen at C1/C14 are not acetate derived but likely derived from molecular oxygen.[72] For cylindrocyclophane D (**1.33**), the lack of the incorporation of L-[methyl-^{13}C]methionine and L-[methyl-^{13}C,^{2}H$_3$]methionine indicated that the branched methyl groups at C2/C15 are not derived from *S*-adenosylmethionine (**1.21**). Instead, the incorporation of the sodium [1,2-^{13}C$_2$]acetate label showed that these methyl groups are also acetate-derived. The authors predicted that the acetate is incorporated in the form of malonate through addition into the carbonyl group in an aldol-type condensation. Subsequent decarboxylation/dehydration and reduction of the resulting exo-methylene would result in the formation of the methyl group.

Based on the results of these feeding experiments, Bobzin and Moore formulated a biosynthetic hypothesis (**Scheme 1.10**). They proposed that a polyketide synthase (PKS) produces a linear nonaketide intermediate (**1.73**), which will be cyclized, decarboxylated and aromatized to form monomeric resorcinol unit (**1.74**). Then, the additional tailoring events, chlorine or methyl group installation, and dimerization lead to the production of nostocyclophane D (**1.60**) and cylindrocyclophane D (**1.33**). In their biosynthetic hypothesis, Bobzin and Moore proposed that the dimerization process occurs between two monomers containing an alkenyl moiety and the final C–C bond formation involves electrophilic substitution of the resorcinol ring. Nevertheless, this biosynthetic hypothesis was very uncertain due to the limited information on the cylindrocyclophane biosynthesis and the unusual nature of the disconnection.



Scheme 1.10: Original biosynthetic hypothesis proposed by Bobzin and Moore involves dimerization of the resorcinol monomers containing an alkenyl group.

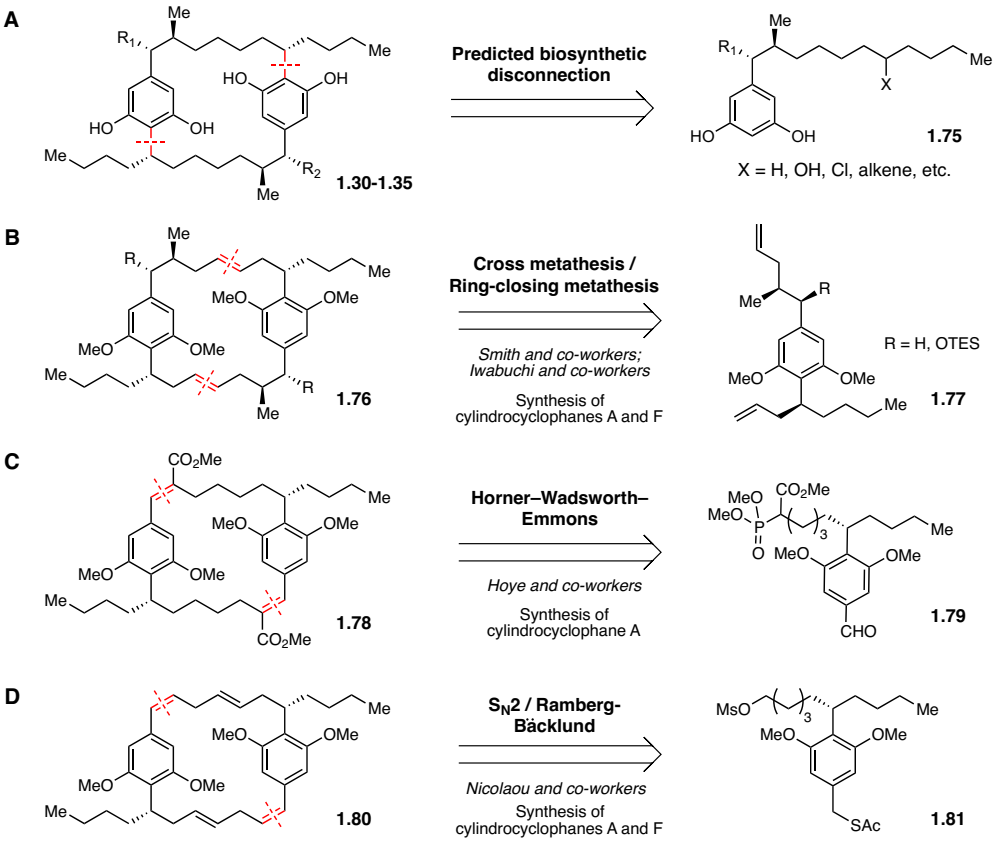
1.3.2. Total syntheses of the cylindrocyclophanes

While the naturally occurring paracyclophanes were not discovered until the early 1990s,[61,62] chemists have been working with paracyclophanes for decades.[73] Since the initial report of the [*m*,*n*]paracyclophanes by Brown and Farthing in 1949,[74] these molecules have been extensively studied in the field of host-guest chemistry and were well known through synthesis.[75] More recently,

paracyclophanes have also been investigated for their use as chiral ligands as well as applications in material science.[76]

Due to their intriguing macrocyclic structure and unique properties, the cylindrocyclophanes have been popular targets for total syntheses. Chemists have developed multiple innovative strategies for assembling the core paracyclophane scaffold. For instance, Smith and co-workers[77] as well as Iwabuchi group[78] employed an elegant tandem cross metathesis–ring-closing metathesis macrocyclization strategy (**Scheme 1.11B**) in their syntheses of (–)-cylindrocyclophanes A (**1.30**) and F (**1.35**). Hoye and co-workers synthesized cylindrocyclophane A (**1.30**) using a double Horner–Wadsworth–Emmons macrocyclization (**Scheme 1.11C**).[79] More recently, Nicolaou and co-workers synthesized (–)-cylindrocyclophanes A (**1.30**) and F (**1.35**) using S_N2 substitution followed by a Ramberg–Bäcklund reaction to construct the paracyclophane macrocycle (**Scheme 1.11D**).[80]

Although numerous approaches have been developed for the total syntheses of the cylindrocyclophanes, it is interesting to note that none of the currently known synthetic routes employ the disconnection predicted to be operational in the biosynthesis of these molecules (**Scheme 1.11**). The use of predicted biosynthetic disconnection requires the C–C bond formation in the dimerization step to be both regio- and stereoselective (**Scheme 1.11A**). Such a transformation is difficult to mimic using currently known synthetic methods. Study of the biosynthesis of cylindrocyclophanes could lead to the discovery of novel enzymes involved in this intriguing dimerization step. Understanding the mechanism by which the enzymes catalyze this transformation may inspire chemists to develop new biomimetic approaches for regio- and stereoselective construction of C–C bonds.



Scheme 1.11: Predicted biosynthetic disconnection and disconnections employed by reported total syntheses of cylindrocyclophanes A and F. **A**) Predicted biosynthetic disconnection. **B**) Disconnection used for the cross metathesis route. **C**) Disconnection used for the Horner–Wadsworth–Emmons route. **D**) Disconnection used for the S_N2-reaction/Ramberg-Bäcklund route.

1.4. Chapter preview

The objective of this thesis project was to discover and characterize the biosynthesis of the cylindrocyclophanes, a family of structurally unique natural products with a rare paracyclophane core. Our main focus was to identify the enzymes required for constructing the conserved [7.7]paracyclophane scaffold. Results of previous feeding experiments and the biosynthetic hypothesis put forth by Bobzin and Moore suggested that the cylindrocyclophane biosynthesis involves highly unusual enzyme chemistry, particularly in the final dimerization step.[72]

In **Chapter 2**, we report the discovery of the cylindrocyclophane (*cyl*) biosynthetic gene cluster through genome mining. Following the discovery of the *cyl* gene cluster, we formulated and validated a

biosynthetic hypothesis for the production of the cylindrocyclophanes through biochemical characterizations of the enzymes encoded by the cluster. In addition, feeding experiments performed in the native producer revealed that decanoic acid is the precursor to the cylindrocyclophanes and the pathway involves C–H activation of the aliphatic carbon center of decanoic acid.

Chapter 3 focuses on the investigation of the termination step of the type I PKS assembly line in the cylindrocyclophane biosynthesis. Through this work, we explored the function of the C-terminal TE domain of the final type I PKS module and discovered that this TE domain has sequence as well as functional similarity to editing thioesterases.

After the initial biochemical characterizations of the enzymes encoded by the *cyl* gene cluster, we investigated the enzymes required for the dimerization step. In **Chapter 4**, we assessed whether the cylindrocyclophane biosynthesis proceeds through direct oxidative dimerization of monomers with unfunctionalized aliphatic carbon center or through pre-functionalization of the aliphatic carbon center followed by dimerization. Our search for the enzyme required for the functionalization of the aliphatic carbon center led to the discovery of novel chlorinating enzymes. Bioinformatic analyses and preliminary biochemical characterizations of these new chlorinases are discussed in detail.

Finally in **Chapter 5**, we report the complete annotation of the enzymes responsible for building the core paracyclophane scaffold of the cylindrocyclophanes. In addition to the new chlorinating enzymes, we also succeeded in identifying a new C–C bond forming enzyme that catalyzes the final dimerization step in the cylindrocyclophane biosynthesis. Biochemical characterizations of the new C–C bond forming enzyme, future mechanistic studies and the potential use of this enzyme as a biocatalyst are discussed.

1.5. References

(1) Bhat, S. V.; Nagasampagi, B. A.; Sivakumar, M. *Chemistry of natural products*; First Edition ed.; Springer: Berlin, **2005**.

(2) Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629.

(3) Medema, M. H.; Fischbach, M. A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11*, 639.

(4) Land, M.; Hauser, L.; Jun, S. R.; Nookaew, I.; Leuze, M. R.; Ahn, T. H.; Karpinets, T.; Lund, O.; Kora, G.; Wassenaar, T.; Poudel, S.; Ussery, D. W. Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* **2015**, *15*, 141.

(5) Streit, W. R.; Schmitz, R. A. Metagenomics—the key to the uncultured microbes. *Curr. Opin. Microbiol.* **2004**, *7*, 492.

(6) Lasken, R. S. Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* **2012**, *10*, 631.

(7) Challis, G. L. Genome mining for novel natural product discovery. *J. Med. Chem.* **2008**, *51*, 2618.

(8) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. Blast+: Architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421.

(9) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339.

(10) Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T. AntiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **2013**, *41*, W204.

(11) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLOS ONE* **2012**, *7*, e34064.

(12) Li, M. H.; Ung, P. M.; Zajkowski, J.; Garneau-Tsodikova, S.; Sherman, D. H. Automated genome mining for natural products. *BMC Bioinformatics* **2009**, *10*, 185.

(13) Gross, H.; Stockwell, V. O.; Henkels, M. D.; Nowak-Thompson, B.; Loper, J. E.; Gerwick, W. H. The genomisotopic approach: A systematic method to isolate products of orphan biosynthetic gene clusters. *Chem. Biol.* **2007**, *14*, 53.

(14) Mohimani, H.; Kersten, R. D.; Liu, W. T.; Wang, M.; Purvine, S. O.; Wu, S.; Brewer, H. M.; Pasa-Tolic, L.; Bandeira, N.; Moore, B. S.; Pevzner, P. A.; Dorrestein, P. C. Automated genome mining of ribosomal peptide natural products. *ACS Chem. Biol.* **2014**, *9*, 1545.

(15) Mohimani, H.; Liu, W. T.; Kersten, R. D.; Moore, B. S.; Dorrestein, P. C.; Pevzner, P. A. NRPquest: Coupling mass spectrometry and genome mining for nonribosomal peptide discovery. *J. Nat. Prod.* **2014**, *77*, 1902.

(16) Soding, J.; Biegert, A.; Lupas, A. N. The HHPred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244.

(17) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLOS ONE* **2009**, *4*, e4345.

(18) Bansal, A. K. Bioinformatics in microbial biotechnology—a mini review. *Microb. Cell Fact.* **2005**, *4*, 19.

(19) de la Torre, M. C.; Sierra, M. A. Comments on recent achievements in biomimetic organic synthesis. *Angew. Chem. Int. Ed.* **2004**, *43*, 160.

(20) Wender, P. A.; Miller, B. L. Synthesis at the molecular frontier. *Nature* **2009**, *460*, 197.

(21) Koeller, K. M.; Wong, C. H. Enzymes for chemical synthesis. *Nature* **2001**, *409*, 232.

(22) Ijaq, J.; Chandrasekharan, M.; Poddar, R.; Bethi, N.; Sundararajan, V. S. Annotation and curation of uncharacterized proteins—challenges. *Front. Genet.* **2015**, *6*, 119.

(23) Erdin, S.; Lisewski, A. M.; Lichtarge, O. Protein function prediction: Towards integration of similarity metrics. *Curr. Opin. Struct. Biol.* **2011**, *21*, 180.

(24) Staunton, J.; Weissman, K. J. Polyketide biosynthesis: A millennium review. *Nat Prod Rep* **2001**, *18*, 380.

(25) Birch, A. J.; Massywestropp, R. A.; Moye, C. J. Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. *Aust. J. Chem.* **1955**, *8*, 539.

(26) Simpson, T. J. Applications of multinuclear NMR to structural and biosynthetic studies of polyketide microbial metabolites. *Chem. Soc. Rev.* **1987**, *16*, 123.

(27) Malpartida, F.; Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature* **1984**, *309*, 462.

(28) Bevitt, D. J.; Cortes, J.; Haydock, S. F.; Leadlay, P. F. 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur. J. Biochem.* **1992**, *204*, 39.

(29) Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **1990**, *348*, 176.

(30) Donadio, S.; Staver, M. J.; Mcalpine, J. B.; Swanson, S. J.; Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **1991**, *252*, 675.

(31) Tuan, J. S.; Weber, J. M.; Staver, M. J.; Leung, J. O.; Donadio, S.; Katz, L. Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene* **1990**, *90*, 21.

(32) Smith, S.; Tsai, S. C. The type I fatty acid and polyketide synthases: A tale of two megasynthases. *Nat. Prod. Rep.* **2007**, *24*, 1041.

(33) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468.

(34) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* **1996**, *3*, 923.

(35) Finking, R.; Solsbacher, J.; Konz, D.; Schobert, M.; Schafer, A.; Jahn, D.; Marahiel, M. A. Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. *J. Biol. Chem.* **2002**, *277*, 50293.

(36) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The phosphopantetheinyl transferases: Catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **2014**, *31*, 61.

(37) Leadlay, P. F. Structural biology: Enzyme assembly line pictured. *Nature* **2014**, *510*, 482.

(38) Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489.

(39) Molnar, I.; Schupp, T.; Ono, M.; Zirkle, R.; Milnamow, M.; Nowak-Thompson, B.; Engel, N.; Toupet, C.; Stratmann, A.; Cyr, D. D.; Gorlach, J.; Mayo, J. M.; Hu, A.; Goff, S.; Schmid, J.; Ligon, J. M. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **2000**, *7*, 97.

(40) Aron, Z. D.; Dorrestein, P. C.; Blackhall, J. R.; Kelleher, N. L.; Walsh, C. T. Characterization of a new tailoring domain in polyketide biogenesis: The amine transferase domain of MycA in the mycosubtilin gene cluster. *J. Am. Chem. Soc.* **2005**, *127*, 14986.

(41) Gu, L. C.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Hakansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459*, 731.

(42) Du, L. C.; Lou, L. L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **2010**, *27*, 255.

(43) Bailey, A. M.; Cox, R. J.; Harley, K.; Lazarus, C. M.; Simpson, T. J.; Skellam, E. Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: First observation of a reductive release mechanism during polyketide biosynthesis. *Chem. Commun.* **2007**, 4053.

(44) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. *Chem. Biol.* **2001**, *8*, 189.

(45) Minto, R. E.; Townsend, C. A. Enzymology and molecular biology of aflatoxin biosynthesis. *Chem. Rev.* **1997**, *97*, 2537.

(46) Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7*, 285.

(47) Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part A—erythromycin biosynthesis). *Nat. Prod. Rep.* **2001**, *18*, 190.

(48) Sthapit, B.; Oh, T. J.; Lamichhane, R.; Liou, K.; Lee, H. C.; Kim, C. G.; Sohng, J. K. Neocarzinostatin naphthoate synthase: An unique iterative type I PKS from neocarzinostatin producer *Streptomyces carzinostaticus*. *FEBS Lett.* **2004**, *566*, 201.

(49) Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **2009**, *48*, 4688.

(50) Chen, H. T.; Du, L. C. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 541.

(51) Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. Type II polyketide synthases: Gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **2007**, *24*, 162.

(52) Mcdaniel, R.; Ebertkhosla, S.; Hopwood, D. A.; Khosla, C. Engineered biosynthesis of novel polyketides - manipulation and analysis of an aromatic polyketide synthase with unproved catalytic specificities. *J. Am. Chem. Soc.* **1993**, *115*, 11671.

(53) Austin, M. B.; Noel, A. J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **2003**, *20*, 79.

(54) Miyanaga, A.; Funa, N.; Awakawa, T.; Horinouchi, S. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 871.

(55) Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in *Bacillus subtilis*. *J. Bacteriol.* **2009**, *191*, 4916.

(56) Fremont, L. Minireview—biological effects of resveratrol. *Life Sci.* **2000**, *66*, 663.

(57) Helfrich, E. J.; Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2015**.

(58) Kwon, H. J.; Smith, W. C.; Xiang, L.; Shen, B. Cloning and heterologous expression of the macrotetrolide biosynthetic gene cluster revealed a novel polyketide synthase that lacks an acyl carrier protein. *J. Am. Chem. Soc.* **2001**, *123*, 3385.

(59) Kwon, H. J.; Smith, W. C.; Scharon, A. J.; Hwang, S. H.; Kurth, M. J.; Shen, B. C–O bond formation by polyketide synthases. *Science* **2002**, *297*, 1327.

(60) Du, L.; Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr. Opin. Drug Discov. Dev.* **2001**, *4*, 215.

(61) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E.; Brinen, L. S.; Kato, Y.; Clardy, J. [7.7]Paracyclophanes from blue-green-algae. *J. Am. Chem. Soc.* **1990**, *112*, 4061.

(62) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E. Structures of cylindrocyclophanes A–F. *Tetrahedron* **1992**, *48*, 3001.

(63) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. Cylindrocyclophanes with proteasome inhibitory activity from the cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2010**, *73*, 1529.

(64) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. Carbamidocyclophanes A–E, chlorinated paracyclophanes with cytotoxic and antibiotic activity from the Vietnamese cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2007**, *70*, 499.

(65) Preisitsch, M.; Harmrolfs, K.; Pham, H. T. L.; Heiden, S. E.; Fuessel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H. J.; Muller, R.; Mundt, S. Anti-MRSA-acting carbamidocyclophanes H–L from the Vietnamese cyanobacterium *Nostoc* sp. CAVN2. *J. Antibiot.* **2015**, *68*, 600.

(66) Luo, S. W.; Kang, H. S.; Krunic, A.; Chlipala, G. E.; Cai, G. P.; Chen, W. L.; Franzblau, S. G.; Swanson, S. M.; Orjala, J. Carbamidocyclophanes F and G with anti-*Mycobacterium tuberculosis* activity from the cultured freshwater cyanobacterium *Nostoc* sp. *Tetrahedron Lett.* **2014**, *55*, 686.

(67) Chen, J. L.; Moore, R. E.; Patterson, G. M. L. Structures of nostocyclophanes A–D. *J. Org. Chem.* **1991**, *56*, 4360.

(68) Kang, H. S.; Santarsiero, B. D.; Kim, H.; Krunic, A.; Shen, Q.; Swanson, S. M.; Chai, H.; Kinghorn, A. D.; Orjala, J. Merocyclophanes A and B, antiproliferative cyclophanes from the cultured terrestrial cyanobacterium *Nostoc* sp. *Phytochemistry* **2012**, *79*, 109.

(69) Eustaquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **2008**, *4*, 69.

(70) Wagner, C.; El Omari, M.; Konig, G. M. Biohalogenation: Nature's way to synthesize halogenated metabolites. *J. Nat. Prod.* **2009**, *72*, 540.

(71) Preisitsch, M.; Heiden, S. E.; Beerbaum, M.; Niedermeyer, T. H.; Schneefeld, M.; Herrmann, J.; Kumpfmuller, J.; Thurmer, A.; Neidhardt, I.; Wiesner, C.; Daniel, R.; Muller, R.; Bange, F. C.; Schmieder, P.; Schweder, T.; Mundt, S. Effects of halide ions on the carbamidocyclophane biosynthesis in *Nostoc* sp. CAVN2. *Mar. Drugs* **2016**, *14*.

(72) Bobzin, S. C.; Moore, R. E. Biosynthetic origin of [7.7]paracyclophanes from cyanobacteria. *Tetrahedron* **1993**, *49*, 7615.

(73) Griffin, R. W. Meta-bridged aromatic compounds. *Chem. Rev.* **1963**, *63*, 45.

(74) Brown, C. J.; Farthing, A. C. Preparation and structure of di-*p*-xylylene. *Nature* **1949**, *164*, 915.

(75) Odashima, K.; Itai, A.; Iitaka, Y.; Koga, K. Host-guest complex formation between a water soluble polyparacyclophane and a hydrophobic guest molecule. *J. Am. Chem. Soc.* **1980**, *102*, 2504.

(76) Hopf, H. [2.2]Paracyclophanes in polymer chemistry and materials science. *Angew. Chem. Int. Ed.* **2008**, *47*, 9808.

(77) Smith, A. B.; Adams, C. M.; Kozmin, S. A.; Paone, D. V. Total synthesis of (–)-cylindrocyclophanes A and F exploiting the reversible nature of the olefin cross metathesis reaction. *J. Am. Chem. Soc.* **2001**, *123*, 5925.

(78) Yamakoshi, H.; Ikarashi, F.; Minami, M.; Shibuya, M.; Sugahara, T.; Kanoh, N.; Ohori, H.; Shibata, H.; Iwabuchi, Y. Syntheses of naturally occurring cytotoxic [7.7]paracyclophanes, (–)-cylindrocyclophane A and its enantiomer, and implications for biological activity. *Org. Biomol. Chem.* **2009**, *7*, 3772.

(79) Hoye, T. R.; Humpal, P. E.; Moon, B. Total synthesis of (–)-cylindrocyclophane a via a double Horner–Emmons macrocyclic dimerization event. *J. Am. Chem. Soc.* **2000**, *122*, 4982.

(80) Nicolaou, K. C.; Sun, Y. P.; Korman, H.; Sarlah, D. Asymmetric total synthesis of cylindrocyclophanes A and F through cyclodimerization and a Ramberg–Bäcklund reaction. *Angew. Chem. Int. Ed.* **2010**, *49*, 5875.

Chapter 2. Discovery and validation of the cylindrocyclophane biosynthetic gene cluster

Parts of this chapter are adopted from previous publications.[1]

2.1. Introduction

The primary goal of this thesis project was to understand how the unusual paracyclophane ring is constructed in the cylindrocyclophane biosynthesis. Through this project, we aimed to discover novel enzymes that are involved in the assembly of the structurally unique [7.7]paracyclophane scaffold, including the enzymes responsible for the stereoselective construction of the sp^2-sp^3 C–C bond linkage during the head-to-tail dimerization reaction of the monomeric resorcinol units. To investigate the enzymatic transformations involved in the cylindrocyclophane biosynthesis, we needed to first identify the biosynthetic gene cluster responsible for the production of these molecules. To begin this project, we sequenced the genome of the producing organism, *Cylindrospermum licheniforme* ATCC 29412 and used a chemically guided genome mining approach to identify a candidate gene cluster.

2.1.1. Enzymes required for the β-methyl incorporation in polyketides

The result of the previously reported feeding experiments in a native cylindrocyclophane A-F producer, *Cylindrospermum licheniforme* ATCC 29204, showed that the β-methyl substituents at the C2/C15 of the cylindrocyclophanes were derived from acetate (**Figure 2.1A**).[2] Because the β-position in polyketides is electrophilic, the methylation process requires a nucleophilic alkyl source (**Figure 2.1B**). While the enzymes necessary for this type of acetate-derived β-branch incorporation in polyketide scaffolds were not known at the time of the feeding experiments, later work on other polyketides has shown that this type of methyl group is incorporated by a highly conserved pathway with homology to the mevalonate pathway in the biosynthesis of isoprenoids.[3]

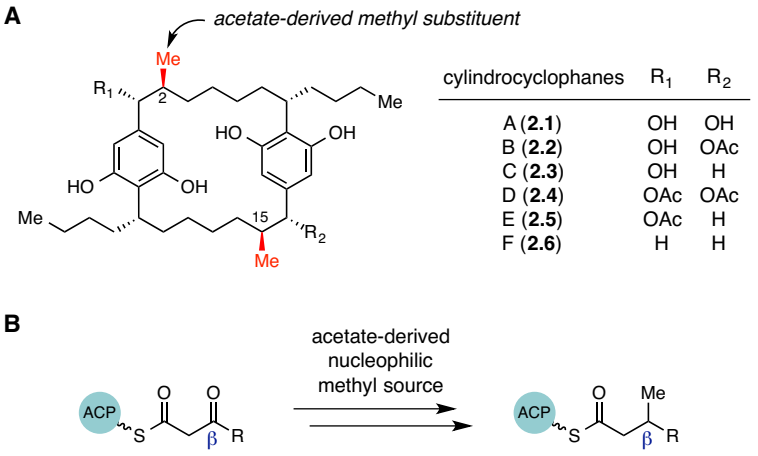
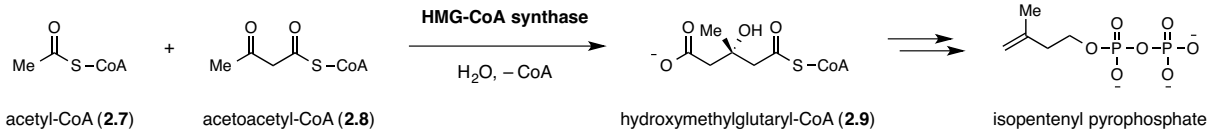


Figure 2.1: The origin of the β-methyl substituents in the cylindrocyclophanes. **A**) Feeding experiment with isotopically labeled sodium acetate revealed that the β-methyl group in the cylindrocyclophanes is derived from acetate.[2] **B**) β-Methylation of polyketides involves an acetate-derived nucleophilic methyl source.

The key enzyme that installs the methyl substituent of isopentenyl pyrophosphate, the key building block used in isoprenoid biosynthesis, is hydroxymethylglutaryl-CoA (HMG-CoA) synthase.[3] HMG-CoA synthase catalyzes the condensation of acetyl-CoA (**2.7**) with acetoacetyl-CoA (**2.8**) to form hydroxymethylglutaryl-CoA (**2.9**, see **Scheme 2.1**). This reaction proceeds through a transfer of an acetyl unit from CoA to a conserved cysteine residue in the HMG-CoA synthase, which then catalyzes the C–C bond formation.[4]

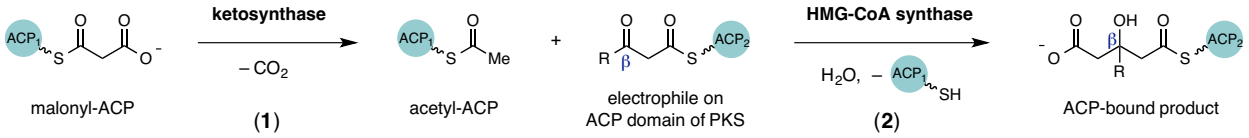


acetyl-CoA (**2.7**) acetoacetyl-CoA (**2.8**) hydroxymethylglutaryl-CoA (**2.9**) isopentenyl pyrophosphate

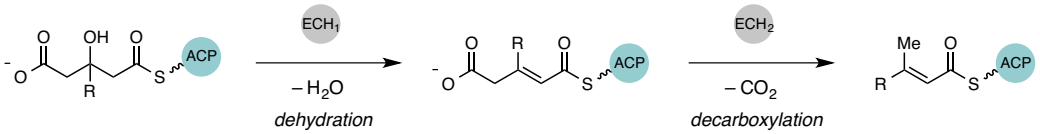
Scheme 2.1: Hydroxymethylglutaryl-CoA formation catalyzed by HMG-CoA synthase in isoprenoid biosynthesis.

Similarly, homologs of HMG-CoA synthase are responsible for the installation of the β-branch in polyketide biosynthetic pathways. One difference between the incorporation of the β-branch in polyketides and isoprenoids is the involvement of acyl carrier protein (ACP) domains. Unlike in the mevalonate pathway, the substrates of the HMG-CoA synthase homolog in polyketide biosynthetic pathways are tethered to ACP domains via thioester linkages (**Scheme 2.2**).[3] The acetyl-ACP substrate for

the HMG-CoA synthase homolog is generated by decarboxylation of a malonyl-ACP thioester. This decarboxylation step is catalyzed by a freestanding ketosynthase (KS) domain that lacks a catalytic cysteine residue (**Reaction 1**, **Scheme 2.2**).[3] Following the β-branch incorporation by the HMG-CoA synthase homolog, the ACP-tethered product undergoes dehydration and decarboxylation to form a β-methyl substituted α,β-unsaturated thioester. The decarboxylation and the dehydration steps are typically catalyzed by two enoyl-CoA hydratase (ECH) domains (**Scheme 2.3**).[3] In the cylindrocyclophane biosynthesis, another domain would be expected to catalyze the reduction of the α,β-unsaturation to complete the installation of the β-methyl group.

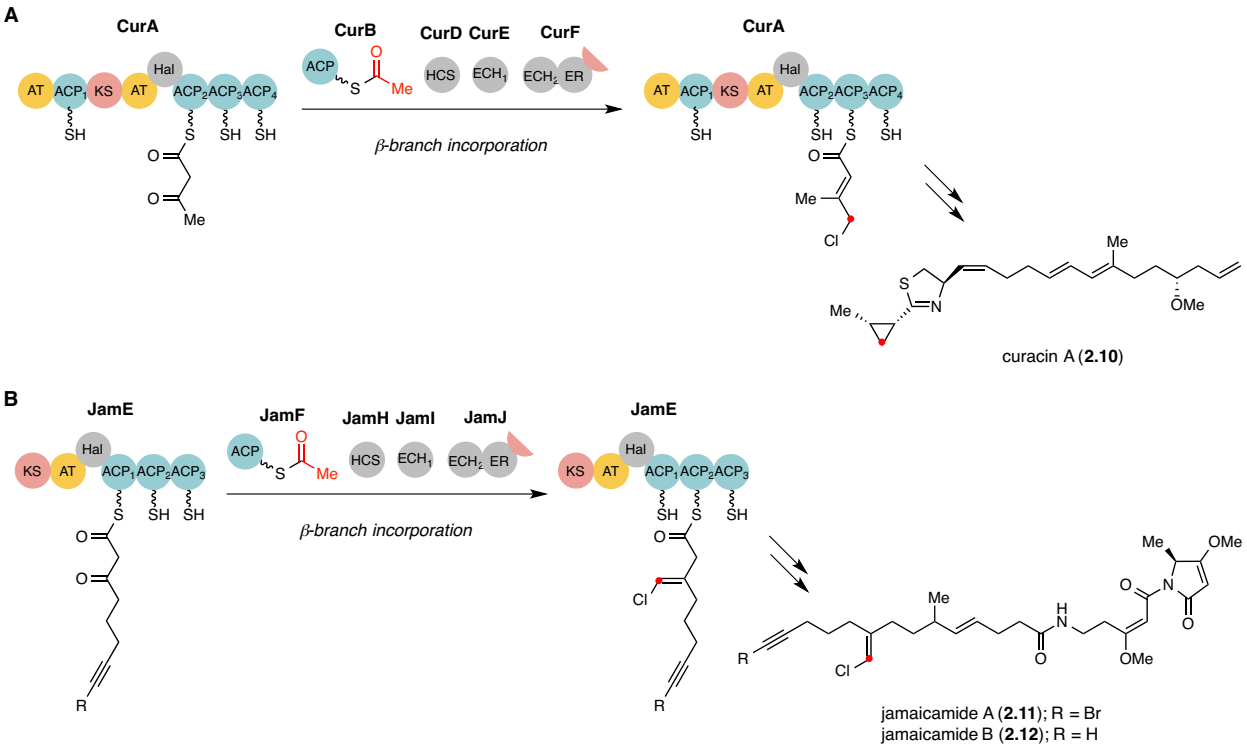


Scheme 2.2: β-branching in polyketides is installed by HMG-CoA synthase homologs that use ACP-bound substrates.



Scheme 2.3: ECH domains catalyze dehydration and decarboxylation to form the β-methyl branch.

Most of the polyketide biosynthetic pathways with the β-branching machinery appear to contain non-standard domain organizations.[5] Examples of polyketides from cyanobacteria with acetate-derived β-branching are the curacins (**Scheme 2.4A**)[6] and the jamaicamides (**Scheme 2.4B**).[7,8] Interesting features found in both curacin and jamaicamide PKS assembly lines include three consecutive ACP domains as well as an unusual halogenase domain. We anticipated that, as in the cases of the curacin and the jamaicamide pathways, the β-branching machinery involved in cylindrocyclophane biosynthesis might employ a unique PKS logic worthy of investigation.



Scheme 2.4: Polyketide biosynthetic pathways in cyanobacteria with the β-branch installation machinery. **A**) Curacin biosynthesis. **B**) Jamaicamide biosynthesis.

2.2. Results and Discussions

2.2.1. Isolation of cylindrocyclophane F from *C. licheniforme* ATCC 29412

Before we started our studies of cylindrocyclophane biosynthesis, we wanted to confirm that the strain we obtained from the American Type Culture Collection (ATCC), *Cylindrospermum licheniforme* ATCC 29412, produced the cylindrocyclophanes in our hands. We also needed to obtain a standard of cylindrocyclophane F (**2.6**) to be used in future studies. To purify cylindrocyclophane F from *C. licheniforme* ATCC 29412, we prepared a 10-L of culture in BG-11 media. The cyanobacterium was cultured at room temperature with constant illumination using a fluorescent light and aeration using an aquarium-pump. After approximately 20 days, the culture of *C. licheniforme* appeared to saturate, and the cells were harvested by centrifugation. The collected cell pellets were lyophilized and cylindrocyclophane

F was purified from the biomass following a previously reported procedure.[9] After organic extraction and several chromatography steps, we were able to obtain 4.6 mg of cylindrocyclophane F (**2.6**).

The structure of the purified cylindrocyclophane F (**2.6**) was confirmed by comparing the ^{1}H and ^{13}C NMR spectra to the previously published data (**Figures 2.2** and **2.3**).[9] In addition, the mass of the purified cylindrocyclophane F (**2.6**) matched that of the expected value by LC-HRMS analysis (**Figure 2.4**). Successful purification of cylindrocyclophane F from *C. licheniforme* ATCC 29412 confirmed that this strain is capable of producing the cylindrocyclophanes under standard culturing conditions.

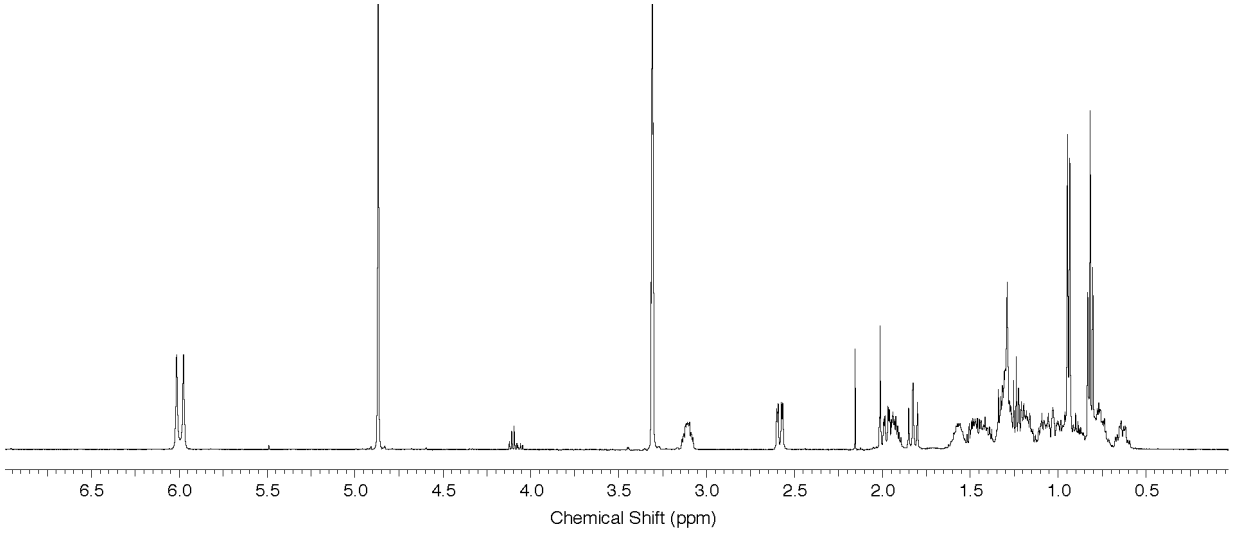


Figure 2.2: ^{1}H-NMR spectrum of cylindrocyclophane F in d_4-methanol (500 MHz).

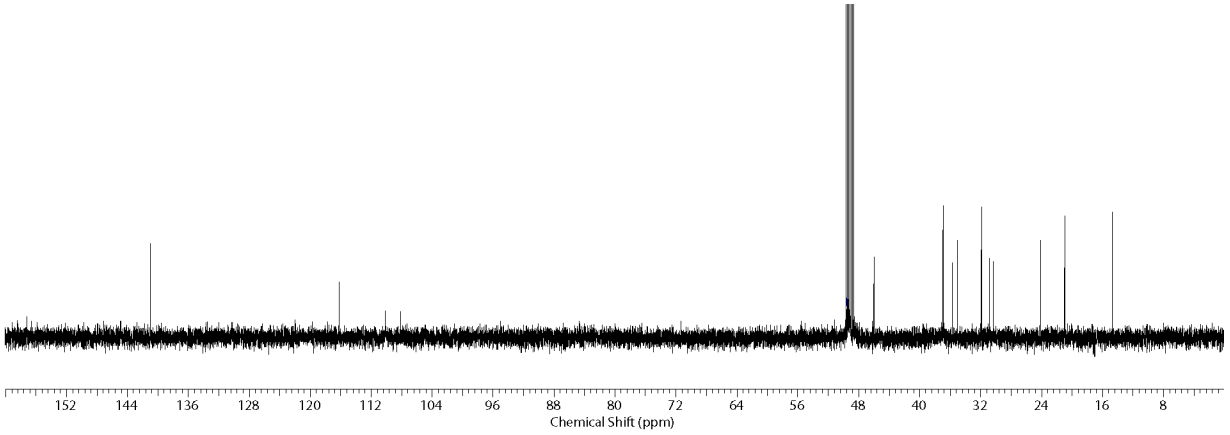


Figure 2.3: ^{13}C-NMR spectrum of cylindrocyclophane F in d_4-methanol (125 MHz).

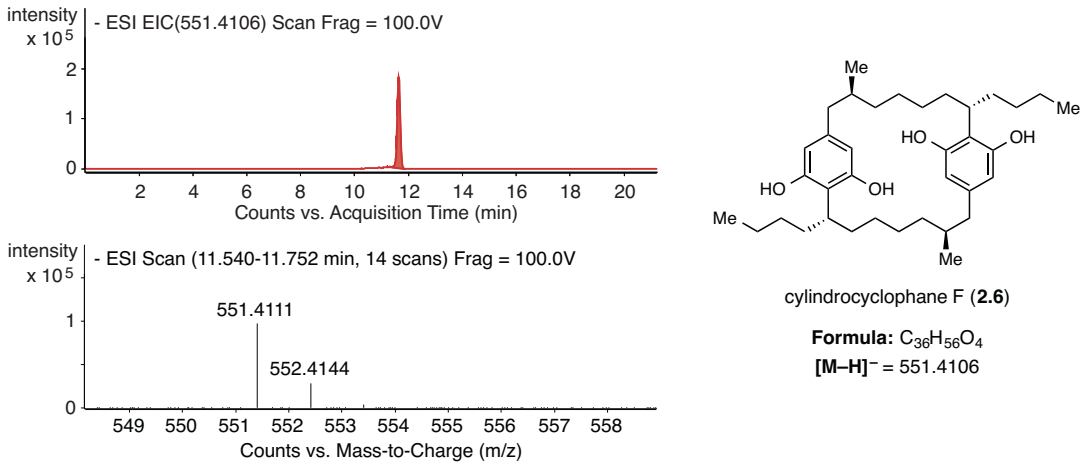


Figure 2.4: LC-HRMS detection of cylindrocyclophane F.

2.2.2. Genome mining for the candidate cylindrocyclophane biosynthetic gene cluster

In order to elucidate the biosynthesis of cylindrocyclophanes, we sought to obtain the sequence of the biosynthetic gene cluster responsible for producing these metabolites. To identify a candidate cylindrocyclophane biosynthetic gene cluster through genome mining, we purified the genomic DNA from a culture of *C. licheniforme* ATCC 29412 and sequenced its genome through shotgun sequencing. The *C. licheniforme* genome sequence data generated by 454 pyrosequencing was converted into a local database.

Prior to obtaining the genome sequencing data, the presence of an HMG-CoA synthase homolog in *C. licheniforme* ATCC 29412 was verified using degenerate PCR using previously reported primer pair.[7] The translated sequence of the amplified PCR product revealed close homology (92 and 90% identity at the amino acid level) to JamH[7] and CurD,[6] which are the HMG-CoA synthase homologs from the jamaicamide and the curacin biosynthesis, respectively.

The nucleotide sequence of the amplified HMG-CoA synthase homolog in *C. licheniforme* was used as a query to perform a BLAST search[10] against the local database containing the genome sequence data of *C. licheniforme.* As a result, a 12.1 kb genome fragment (contig 1024) containing a full-length HMG-CoA synthase homolog was identified. This genome fragment also contained six complete (*cylA–cylF*) and one

partial (*cylG*) open reading frames (ORFs) that appeared to encode enzymes with functions consistent with the cylindrocyclophane biosynthesis, including the other enzymes required for the α-methyl installation (*cylE–G*) and a type I PKS (*cylD*). Two other contigs (111 and 112) contained fragments of the ECH domain (CurE and CurF homologs) and a type I PKS (*cylH*) that might be involved in the cylindrocyclophane biosynthesis. The sequences of the cylindrocyclophane gene cluster on contigs 1024, 111, and 112 were assembled into one continuous DNA sequence by PCR amplification of the missing regions from *C. licheniforme* genomic DNA, cloning of PCR products, and sequencing. The candidate cylindrocyclophane (*cyl*) biosynthetic gene cluster discovered through initial genome sequencing and genome mining contained 12 ORFs (*cylA-L*). The work done in this chapter was done based on this initial *cyl* gene cluster sequence.

While we identified a promising candidate *cyl* gene cluster, we were uncertain whether the gene cluster we found was complete. While one cluster boundary was fairly well-determined, the other cluster boundary could not be confirmed since we could not predict which contigs might contain other enzymes that might be necessary for the biosynthesis. To determine if there are additional ORFs that may be involved in the cylindrocyclophane biosynthesis, we re-sequenced the genome of *C. licheniforme* ATCC 29412. The second genome sequencing and assembly were performed at Cofactor Genomics using Illumina sequencing methods, and we found a 957.9 kb genome fragment (scaffold 7) that contained the entire candidate *cyl* gene cluster. The cluster boundaries were determined based on the presence of possible housekeeping genes such as histidine kinase on one boundary and a transposase on the other boundary. In addition to the 12 ORFs that we had previously found, there are four more ORFs (*cylM-P*) that are predicted to have a function in the cylindrocyclophane biosynthesis (**Figure 2.5**). The 16 ORFs in the candidate *cyl* gene cluster were annotated based on the conserved domain analysis on the NCBI website,[11] and the domain organizations of the PKSs were annotated using the PKS/NRPS Analysis online tool (**Table 2.1**).[12] As predicted, the candidate *cyl* gene cluster contains genes encoding all of the enzymatic machinery required for the incorporation of the α-methyl substituent (*cylE, F, G*, *H*). There are

also two type I modular PKS assembly line enzymes (*cylD*, *H*) as well as a type III PKS (*cylI*). Two genes are predicted to be involved in the fatty acid activation (*cylA*, *B*), and the remaining genes are predicted to be responsible for tailoring reactions (*cylC, cylJ-P*).

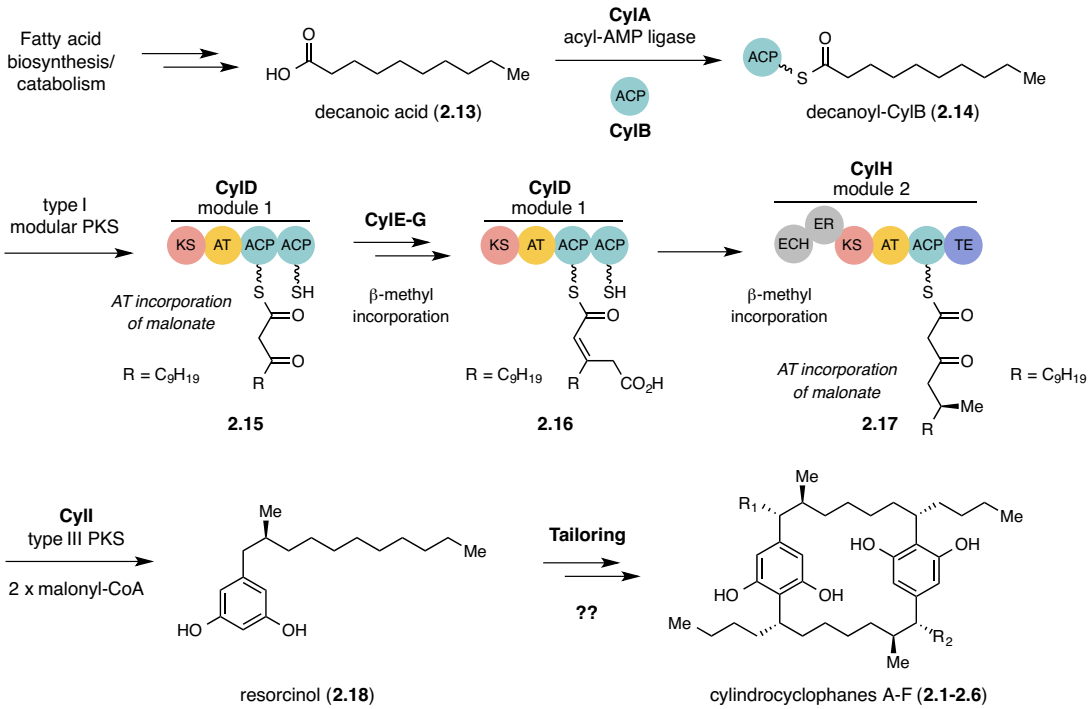
Figure 2.5: ORF diagram of the candidate *cyl* gene cluster.

Table 2.1: Annotated functions of the ORFs in the *cyl* gene cluster.

Protein	Size, aa	Proposed function	Protein homolog/strain	Accession number	Identity/ similarity %
CylA	605	AMP-dependent synthase and ligase	Cyan8802_3741, *Cyanothece* sp. PCC 8802	ACV02552	62/78
CylB	125	Acyl carrier protein	gsl1945, *Gloeobacter violaceus* PCC 7421	BAC89886	52/72
CylC	471	Hypothetical protein	Cyan7822_5858, *Cyanothece* sp. PCC 7822	ADN17712	51/69
CylD	1228	PKS (KS-AT-ACP-ACP)	Npun_F3359, *Nostoc punctiforme* PCC 73102	ACC81786	59/73
CylE	399	Ketosynthase (□-methyl group installation)	CurC, *Lyngbya majuscula* 3L	AAT70098	73/86
CylF	419	HMG-CoA synthase (□-methyl group installation)	CurD, *Lyngbya majuscula* 3L	AEE88286	87/95
CylG	254	Enoyl CoA Hydratase/Isomerase	CurE, *Lyngbya majuscula* 3L	AAT70100	81/90
CylH	2232	PKS (ECH-ER-KS-AT-ACP-TE)	CurE, *Lyngbya majuscula* 3L	AEE88284	62/76
CylI	373	Type III polyketide synthase	PCC7424_2733, *Cyanothece* sp. PCC 7424	ACK71145	60/79
CylJ	181	Methyltransferase	PCC7424_2732, *Cyanothece* sp. PCC 7424	ACK71144	61/80
CylK	651	Hemolysin-type calcium-binding protein	FJSC11DRAFT_3960, *Fischerella* sp. JSC-11	EHC09323	45/63
CylL	535	Flavin-dependent oxidoreductase	Th970_020100003472, *Thiorhodovibrio* sp. 970	ZP_09809663	36/57
CylM	232	Aldo/keto reductase	*Mycobacterium xenopi* Rvm700367	ZP_00979622	35/79
CylN	435	Glycosyltransferase_	*Calothrix sp.* PCC 7507	YP_007065913	59/97
CylO	448	Transferase	*Microcystis aeruginosa* PCC 9443	WP_002766670	58/77
CylP	514	Rieske enzyme	*Fischerella* sp. PCC 9605	WP_026733903	62/75

2.2.3. Biosynthetic hypothesis for the cylindrocyclophane assembly

Based on the detailed annotations of the candidate *cyl* gene cluster, we formulated a biosynthetic hypothesis for the cylindrocyclophane assembly (**Scheme 2.5**). An unexpected feature of the candidate *cyl* gene cluster was the abbreviated nature of the type I modular PKS assembly line. CylD and CylH each contain only one intact module, which suggests that this type I PKS assembly line only catalyzes two elongation reactions. It would be unlikely for CylD and CylH to function iteratively to generate the entire alkyl chain of the cylindrocyclophanes, because both enzymes lack the domains needed to form a fully reduced polyketide backbone. Both CylD and CylH acyltransferase (AT) domains are expected to activate malonyl-CoA as extender unit based on the presence of the conserved HAFH motif associated with malonyl-CoA incorporation (**Figure 2.6**).[13]



Scheme 2.5: Biosynthetic hypothesis for the production of cylindrocyclophane in *C. licheniforme*.

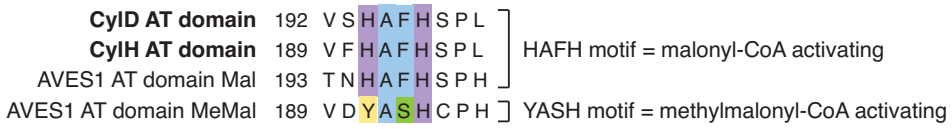
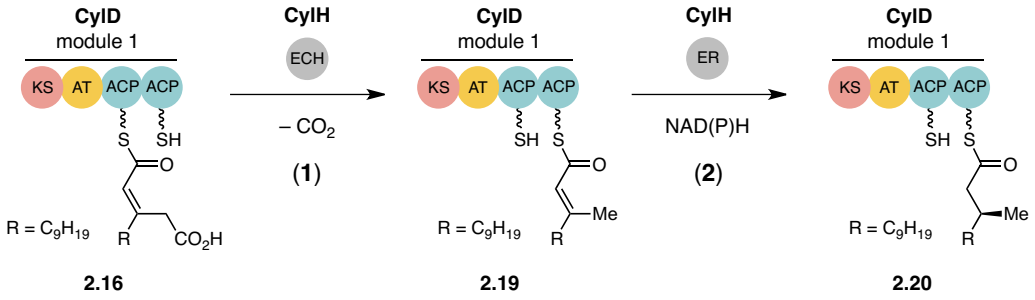


Figure 2.6: Malonyl-CoA activation motif in the CylD and CylH AT domains.

Based on the domains present in the type I PKS assembly line and the structure of the cylindrocyclophanes, we proposed that the biosynthesis of these metabolites originated from decanoic acid (**2.13**), which could be derived from either fatty acid biosynthesis or catabolism (**Scheme 2.5**). Activation of decanoic acid (**2.13**) by the fatty acyl-AMP ligase (CylA) and loading onto the freestanding ACP (CylB) would lead to the formation of decanoyl-CylB (**2.14**). The activated decanoyl thioester could then be transferred to the first type I PKS (CylD), which has KS, AT and two tandem ACP domains. The presence of extra ACP domains is a common feature of the □-branch installation machinery.[3] CylD is expected to perform one elongation with malonyl-CoA to form a □-ketothioester (**2.15**). This ACP-tethered intermediate could then be processed by the set of enzymes that incorporates the □-methyl substituent (CylE, F and G) to generate the □,□-unsaturated thioester (**2.16**). The ECH domain of CylH could then catalyze decarboxylation (**Reaction 1**, **Scheme 2.6**), and the enoylreductase (ER) domain of CylH could stereoselectively reduce the □,□-unsaturated thioester intermediate (**2.19**) using NAD(P)H as cofactor (**Reaction 2**, **Scheme 2.6**). We predict that processing by the ECH and the ER domains of CylH occurs on intermediates tethered to the second ACP domain of CylD. The final module of CylH that contains KS, AT, ACP and TE domains likely catalyzes a second elongation with malonyl-CoA to form the polyketide assembly line product (**2.17**, see **Scheme 2.5**).



Scheme 2.6: Decarboxylation and reduction catalyzed by the ECH and the ER domains of CylH.

Next, as depicted in **Scheme 2.5**, we proposed that the type III PKS CylI directly used the CylH-bound □-ketothioester (**2.17**) as a substrate, terminating the type I PKS assembly line in place of the CylH TE

domain. Type III PKSs typically use CoA thioesters as substrates,[14] and this type of direct interaction between a type I modular PKS and a type III PKS has not been reported previously. CylI could catalyze two successive Claisen condensations with two units of malonyl-CoA, cyclization, and aromatization to form the resorcinol-type aromatic ring. The resulting resorcinol derivative (**2.18**) could be a potential monomeric precursor to the cylindrocyclophanes. We predicted that the other tailoring enzymes (CylC, CylJ-P) could be responsible for generating the cylindrocyclophanes (**2.1**-**2.6**) from this resorcinol intermediate (**2.18**). Based on our limited knowledge on the other tailoring enzymes encoded by the *cyl* gene cluster, we could not easily predict how the elaboration of the paracyclophane would occur. Overall, the PKS machinery encoded by the *cyl* gene cluster is consistent with its potential role in the cylindrocyclophane biosynthesis.

2.2.4. Biochemical characterization of the decanoic acid activation by CylA and CylB

To link the putative *cyl* gene cluster to the cylindrocyclophane biosynthesis, we decided to study the *in vitro* chemical functions of the enzymes in the pathway. We were unable to rely on gene inactivation in the native producer to connect the *cyl* gene cluster to the cylindrocyclophane production, because *C. licheniforme* is not genetically tractable. Instead, the target enzymes were cloned into expression vectors for heterologous production and the functions of the purified enzymes were characterized *in vitro*.

We first examined fatty acid activation catalyzed by the fatty acyl-AMP ligase CylA and the ACP CylB. Through hetereologous expression in *Escherichia coli* (*E. coli*), we successfully obtained N-His_6 tagged CylA and C-His_6 tagged CylB (**Figure 2.7**). Purified C-His_6-CylB was confirmed to be in the *apo* form (**2.21**) using the the BODIPY-CoA (**2.22**)[15] fluorescent phosphopantetheinylation assay. In this assay, the promiscuous phosphopantetheinyl transferase Sfp[16] is used to install a fluorophore-tagged CoA analog (**2.22**) onto ACP domains (**Scheme 2.7** and **Figure 2.8**).

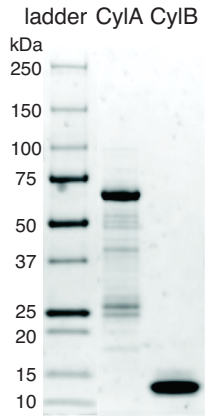
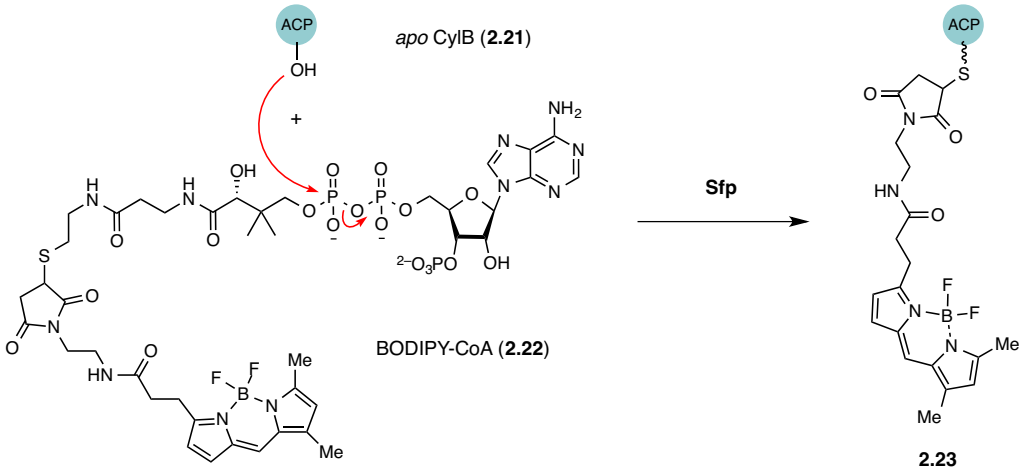


Figure 2.7: SDS-PAGE of purified His_6-tagged CylA and CylB.



Scheme 2.7: BODIPY-CoA loading of *apo* CylB by Sfp.

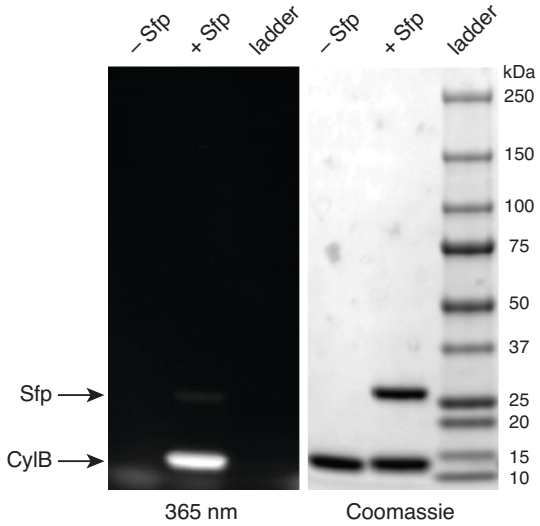
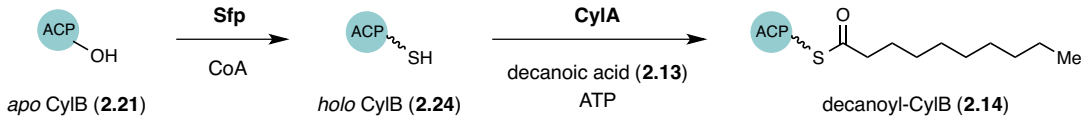


Figure 2.8: SDS-PAGE analysis of the BODIPY-CoA fluorescent phosphopantetheinylation assay. CylB is post-translationally modified with BODIPY-CoA only in the presence of Sfp.

Next, we tested the ability of CylA to activate decanoic acid using ATP in *in vitro* assays (**Scheme 2.8**). The *holo* CylB required for loading of the CylA-activated decanoic acid was generated using Sfp and CoA.[16] Upon incubation of decanoic acid (**2.13**) with the *holo* CylB (**2.24**) in the presence of CylA, a peak corresponding to decanoyl-CylB (**2.14**) was observed by HPLC (**Figure 2.9**). Decanoyl-CylB formation was not observed in absence of any of the active assay components. The assay result confirmed that CylA catalyzes activation of decanoic acid (**2.13**) for loading onto CylB as we predicted.



Scheme 2.8: Decanoic acid loading onto *holo* CylB catalyzed by CylA.

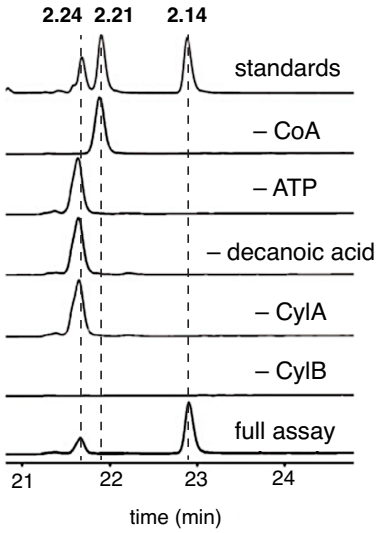
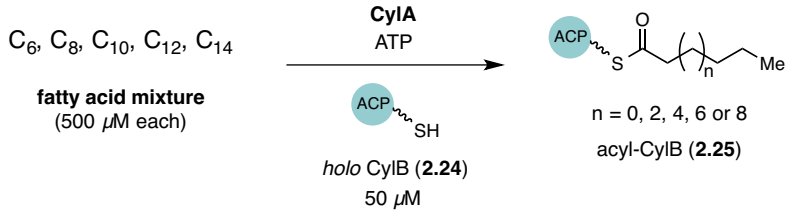


Figure 2.9: HPLC analysis of the CylA and CylB assay to generate decanoyl-CylB (monitored at 220 nm).

To determine whether CylA was also capable of activating fatty acids with different chain lengths, we performed an LC-MS-based competition assay. In this assay, equal amounts of C_6, C_8, C_{10}, C_{12} and C_{14} saturated fatty acids were incubated with CylA and *holo* CylB (**2.24**), and the resulting acyl-CylB (**2.25**) products were analyzed by LC-MS (**Scheme 2.9**). The competition assay revealed that CylA selectively

activates decanoic acid over fatty acids of other chain lengths (**Figure 2.10**). We observed a small amount (less than 10%) of octanoyl-CylB, but we did not detect any of the other possible acyl-CylB masses.



Scheme 2.9: Fatty acid competition assay to determine the selectivity of CylA.

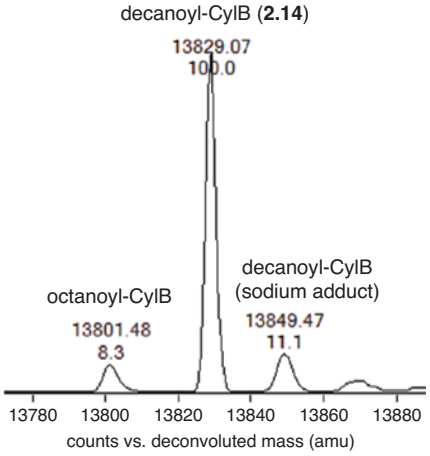
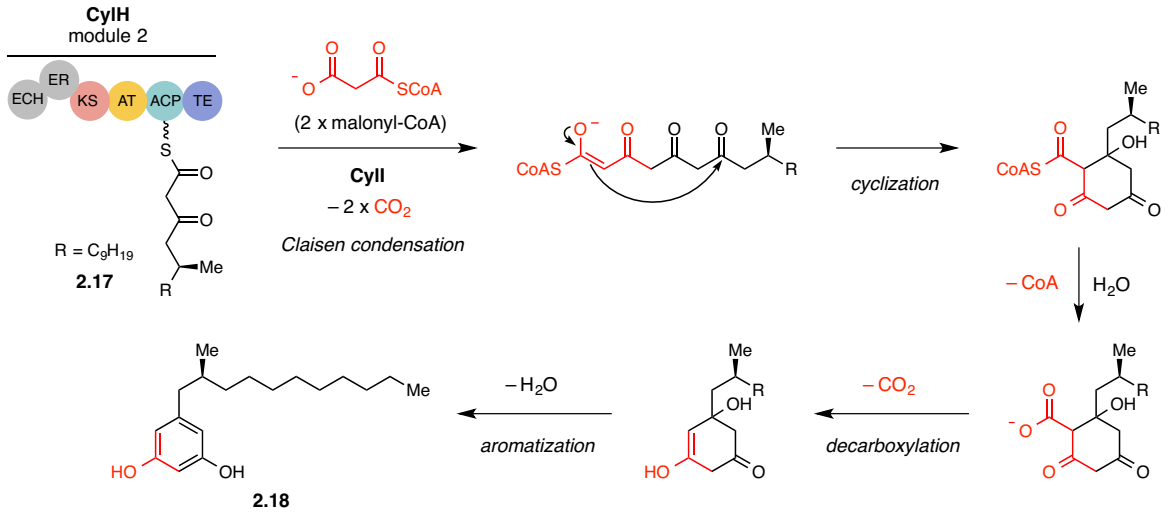


Figure 2.10: LC-MS analysis of detection of the acyl-CylB in the fatty acid competition assay.

The results of the biochemical characterization of CylA and CylB support our biosynthetic hypothesis. The observed selectivity of CylA for decanoic acid (**2.13**) is consistent with the observation that all members of the cylindrocyclophanes and their analogs share the same core paracyclophane scaffold.[9,17-21] The use of a free fatty acid as a starter unit is uncommon among type I modular PKS assembly lines, but has been demonstrated previously in the biosynthesis of mycobacterial fatty acids[22,23] and several hybrid PKS/non-ribosomal peptide synthetase (NRPS) pathways.[7,24]

2.2.5. Biochemical characterization of the CylI-catalyzed resorcinol formation

Another important enzyme to characterize in order to test for our biosynthetic hypothesis was the type III PKS CylI. We proposed that CylI would use a CylH-bound β-ketothioester substrate (**2.17**) to form the resorcinol moiety (**2.18**) through two cycles of Claisen condensation using malonyl-CoA as extender unit, followed by cyclization, thioester hydrolysis, decarboxylation and aromatization (**Scheme 2.10**).



Scheme 2.10: Proposed mechanism for resorcinol ring formation by CylI.

As for CylA and CylB, we cloned *cylI* into an expression vector and heterologously expressed the encoded enzyme in *E. coli* to purify the N-His_6 tagged CylI (**Figure 2.11**). In addition to the enzyme, we also needed to access the predicted substrate for CylI. To study the activity of CylI, we therefore synthesized the *N*-acetylcysteinamine thioester (SNAC) analog (**2.26**, see **Figure 2.12**) of the predicted CylI substrate, the CylH ACP-tethered β-ketothioester (**2.17**), as a racemic mixture.

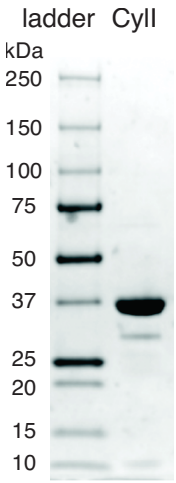


Figure 2.11: SDS-PAGE of purified the N-His_6-tagged CylI.

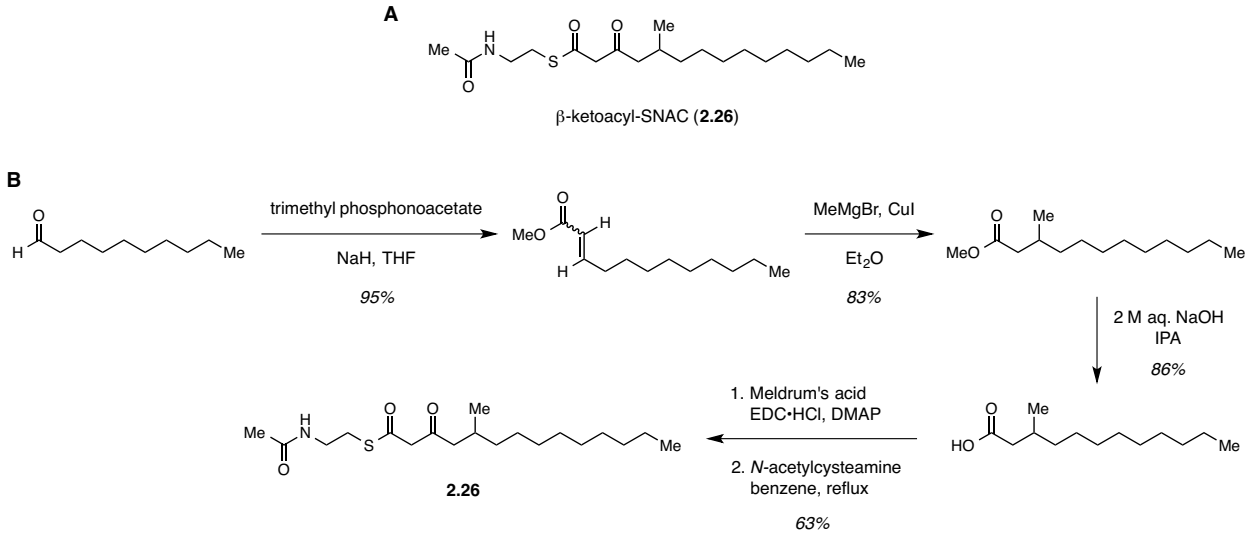
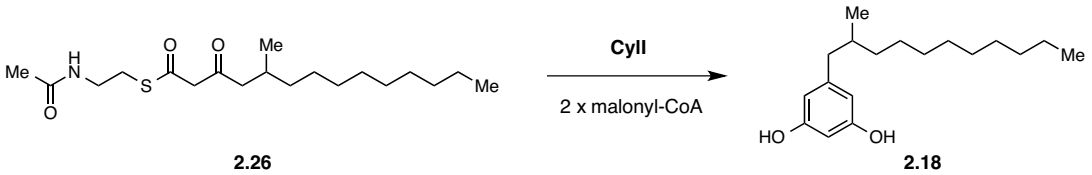


Figure 2.12: The synthesis of the CylH ACP-tethered β-ketothioester (**2.17**) analog for CylI assays. **A**) Structure of the β-ketoacyl-SNAC substrate **2.26**. **B**) Synthetic route for the β-ketoacyl-SNAC substrate **2.26**.

The activity of CylI was assayed *in vitro* by incubating the synthetic β-ketoacyl-SNAC substrate (**2.26**) in the presence of CylI and malonyl-CoA (**Scheme 2.11**). We observed a single new product by HPLC analysis of the assay mixture (**Figure 2.13A**). This peak was not observed when any of the active assay components were left out (**Figure 2.13B**).



Scheme 2.11: *In vitro* CylI assay using synthetic SNAC substrate.

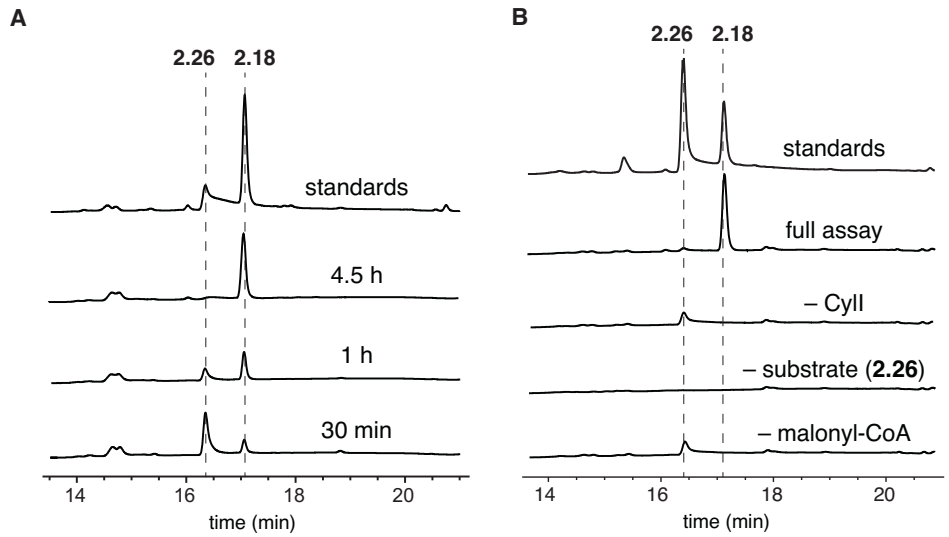


Figure 2.13: HPLC analysis of the CylI assay (monitored at 230 nm). **A**) Time course of the full assay. **B**) Comparison of the full assay and the negative controls.

To determine the structure of the newly formed product, we isolated it from a large-scale enzymatic reaction. In addition, we prepared a synthetic standard of the predicted resorcinol product and compared its ^{1}H and ^{13}C NMR spectra (**Figure 2.14** and **2.15**) and LC-HRMS spectrum (**Figure 2.16**) to those of the enzymatic product (**2.18**). While the ^{1}H NMR spectra of the CylI assay product was slightly shifted compared to the synthetic standard, this could potentially arise from impurities present in the CylI assay product. The ^{1}H NMR spectrum of the synthetic resorcinol standard also shows a peak corresponding to the phenol hydrogen at 6.08 ppm, which is the absent in the ^{1}H NMR spectrum of the CylI assay product. ^{13}C NMR spectra and the LC-HRMS spectra matched perfectly. The CylI-catalyzed resorcinol formation was determined to have slight enantioselectivity (*ee* = 20%) by chiral HPLC analysis (**Figure 2.17**). Based on the NMR and the HRMS characterizations, we determined that the product purified from the enzymatic assay is identical to the product standard prepared by chemical synthesis.

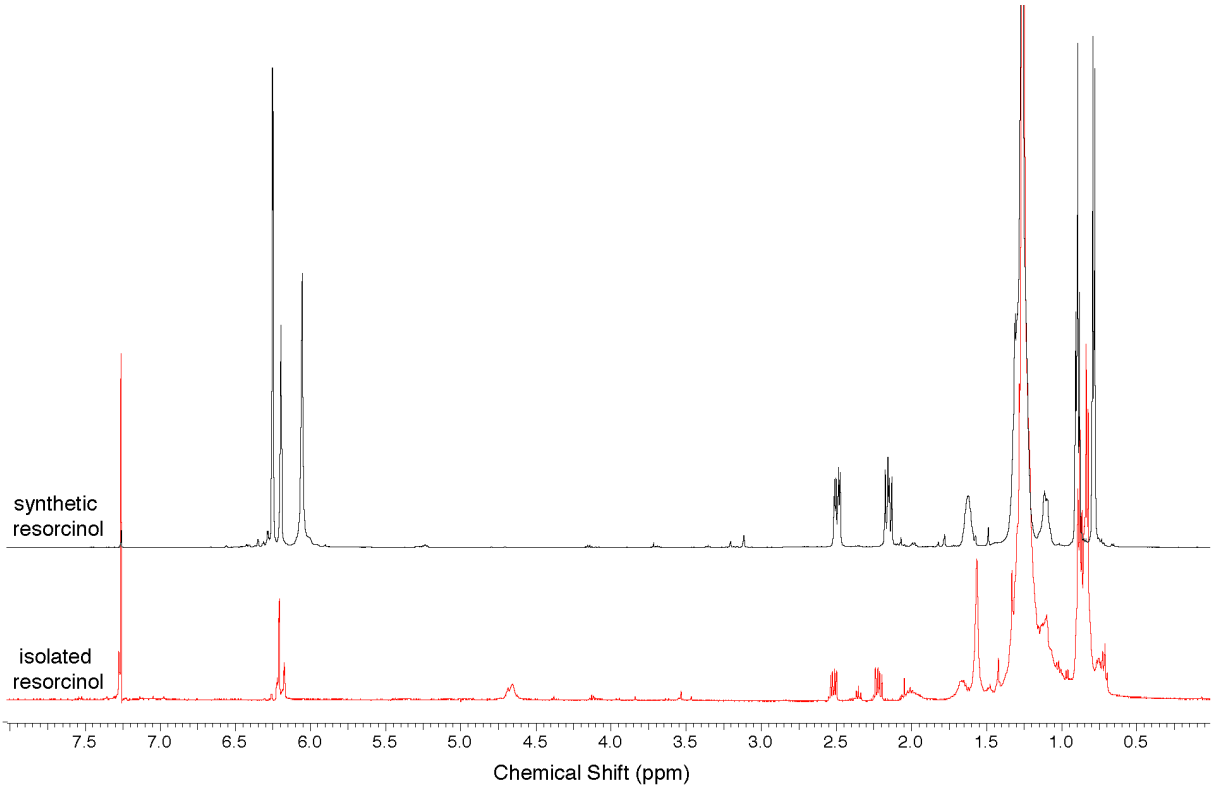


Figure 2.14: Overlay of ^{1}H NMR spectra of the synthetic and the isolated resorcinol in *d*-chloroform (500 MHz).

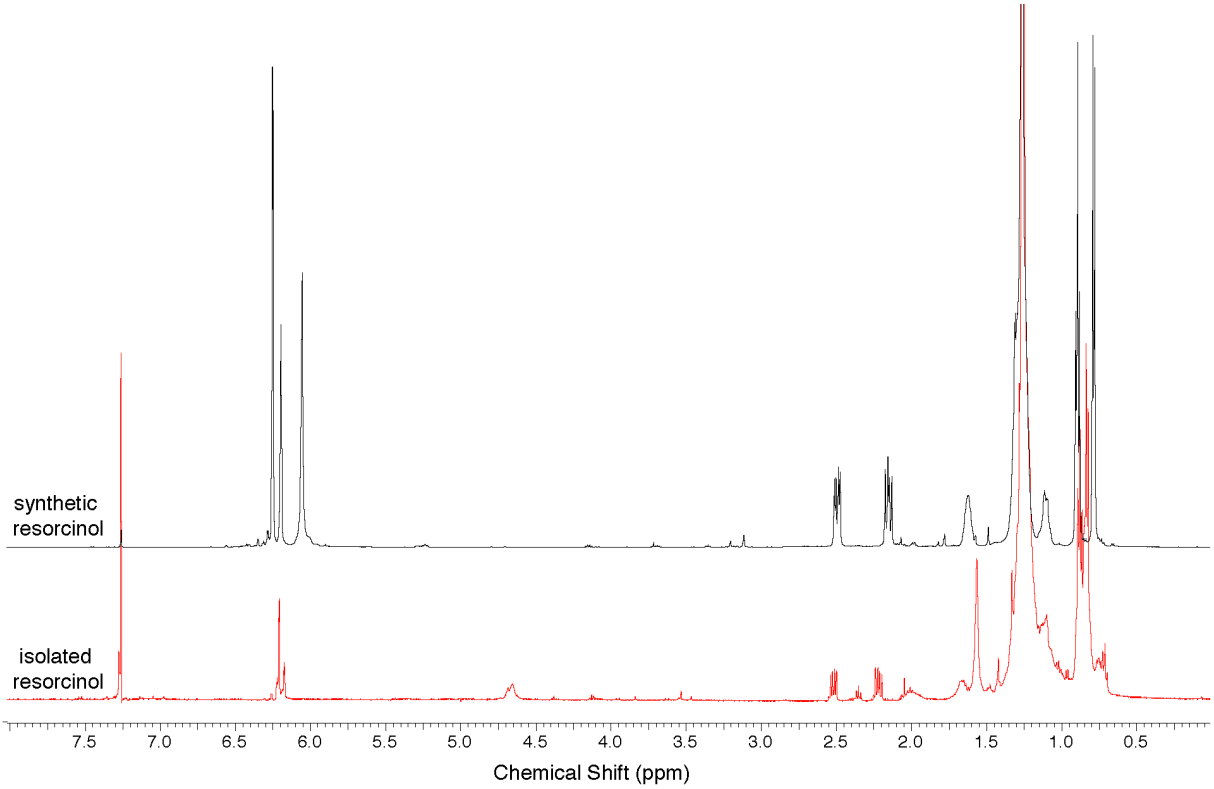


Figure 2.15: Overlay of ^{13}C NMR spectra of the synthetic and the isolated resorcinol in *d*-chloroform (125 MHz).

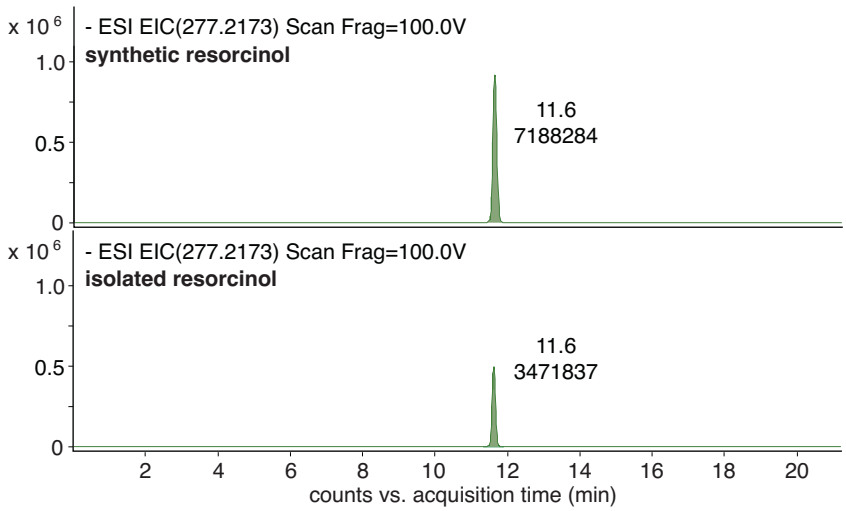


Figure 2.16: LC-HRMS extracted ion chromatograms of the synthetic and the isolated resorcinol (**2.18**).

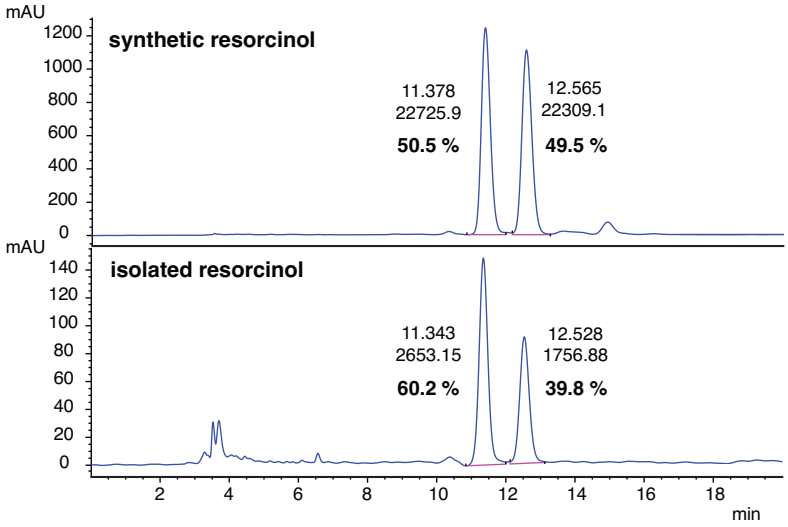
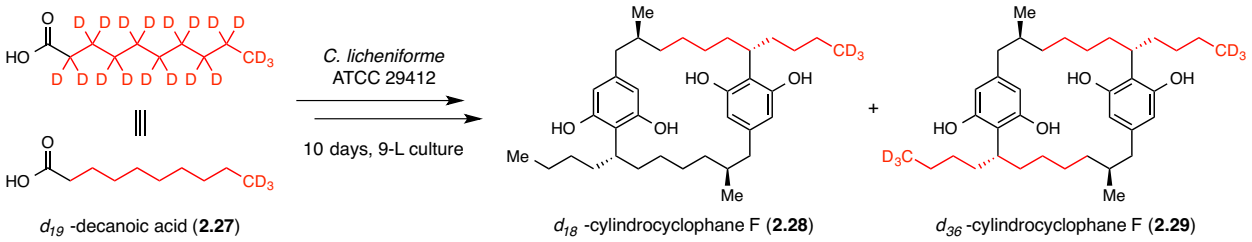


Figure 2.17: Chiral HPLC analysis of the synthetic and the isolate resorcinol (monitored at 210 nm).

The results of CylI assay therefore supported our hypothesis that this type III PKS uses a substrate that is covalently tethered to the type I PKS CylH. Similar use of acyl-ACP substrates by several type III PKS enzymes and domains have been reported, but the acyl-ACP substrates are generated by iterative type I[25,26] and type II[27-29] fatty acid synthases in these previous cases. To our knowledge, CylI is the first characterized type III PKS to accept an acyl-ACP substrate generated by a type I modular PKS assembly line.

2.2.6. Feeding study links the activation of decanoic acid to the cylindrocyclophane biosynthesis

The *in vitro* activities of CylA, CylB and CylI supported our biosynthetic hypothesis for the cylindrocyclophane assembly and the involvement of the *cyl* gene cluster in the cylindrocyclophane biosynthesis. To further confirm that the fatty acid activation pathway encoded by this cluster was directly connected to the cylindrocyclophane production, we performed feeding studies with *C. licheniforme* ATCC 29412. We evaluated the incorporation of the isotopically labeled precursor, d_{19}-decanoic acid, into the cylindrocyclophane scaffold using LC-HRMS of the culture extracts (**Scheme 2.12**).



Scheme 2.12: Incorporation of d_{19}-decanoic acid into cylindrocyclophane F in the feeding experiment.

The organic extracts of a 9-L *C. licheniforme* culture fed with 100 mg (65 µM) of d_{19}-decanoic acid (**2.27**) contained masses corresponding to both d_{18}- (**2.28**) and d_{36}-cylindrocyclophane F (**2.29**, see **Figure 2.18**). The masses corresponding to these single and double incorporation products were not found in the extracts from cultures fed with unlabeled decanoic acid (**2.18**). Deuterium-label incorporation was also observed for cylindrocyclophanes A-E (**2.1**-**2.5**). This result verifies that decanoic acid is a biosynthetic precursor to the cylindrocyclophanes and is incorporated into both halves of the natural product carbon skeleton. The integration of decanoic acid into the cylindrocyclophanes also implies that the eventual C7/C20 carbon atoms of the paracyclophane scaffold enter the biosynthetic pathway as an unactivated methylene group (**Scheme 2.13**). Construction of this unique macrocyclic scaffold therefore requires a C–H functionalization step at some point in the biosynthetic pathway.

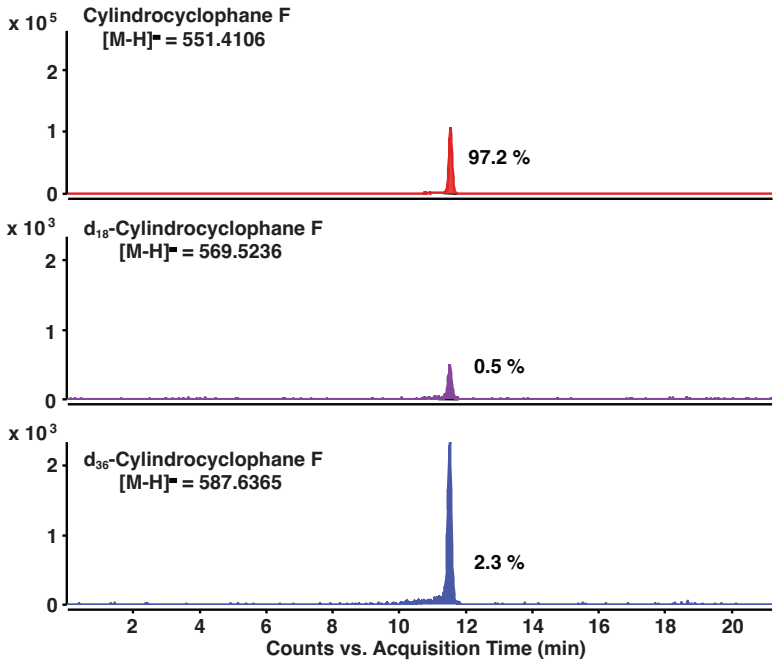
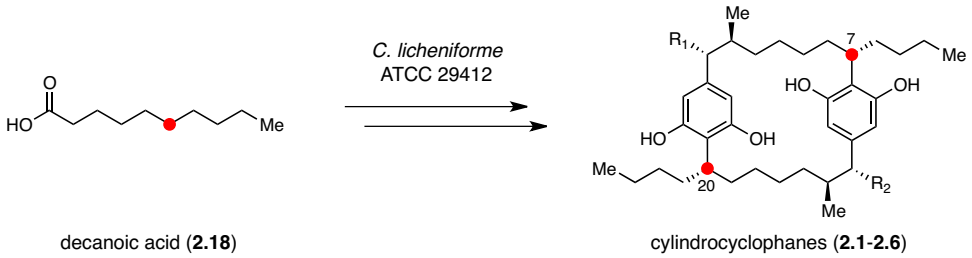


Figure 2.18: LC-HRMS extracted ion chromatograms of unlabeled and deuterium-labeled cylindrocyclophane F in d_{19}-decanoic acid feeding experiment.



Scheme 2.13: C7/C20 carbon atoms of the cylindrocyclophanes correspond to an unactivated carbon atom in the decanoic acid biosynthetic precursor.

Unexpectedly, we observed a higher abundance of d_{36}-cylindrocyclophane F (**2.29**, 2.3% of cylindrocyclophane F) than d_{18}-cylindrocyclophane F (**2.28**, 0.5% of cylindrocyclophane F) in the culture extract, despite the low levels of incorporation. This finding might be due to the higher abundance of the d_{19}-decanoic acid (**2.27**) compared to the free decanoic acid in the culture of *C. licheniforme* at the time of feeding. Because the addition of d_{19}-decanoic acid to the cultures at the early stages inhibits the growth of *C. licheniforme*, d_{19}-decanoic acid was added to the cultures after 10 days. The unlabeled cylindrocyclophane F (**2.6**) may have accumulated prior to the addition of d_{19}-decanoic acid. Once d_{19}-

decanoic acid is added, the deuterium-labeled precursor could outcompete the unlabeled precursor for incorporation into the pathway. In addition, this observation could also suggest that monomer generation and paracyclophane formation are tightly coupled, and the resorcinol intermediate does not accumulate. We also searched for other possible pathway intermediates with various functional group modifications (hydroxylation, halogenation, dehydrogenation etc.), but these masses were not detected.

Analysis of the isotope distribution of d_{19}-decanoic acid precursor (**Table 2.2** and **Figure 2.19A**) and the deuterium-labeled cylindrocyclophane F (**Tables 2.3** and **2.4**, **Figure 2.19B**) revealed that only one deuterium label is lost in each molecule of d_{19}-decanoic acid incorporated into the cylindrocyclophane scaffold. If the cylindrocyclophane biosynthesis proceeds through an alkene intermediate as originally hypothesized by Bobzin and Moore,[2] we would expect to see loss of two deuterium labels. Based on this result, we hypothesized that cylindrocyclophane biosynthesis involves a radical-mediated C–H functionalization step catalyzed by one or more of the remaining enzymes encoded by the *cyl* cluster (CylC, CylJ-P).

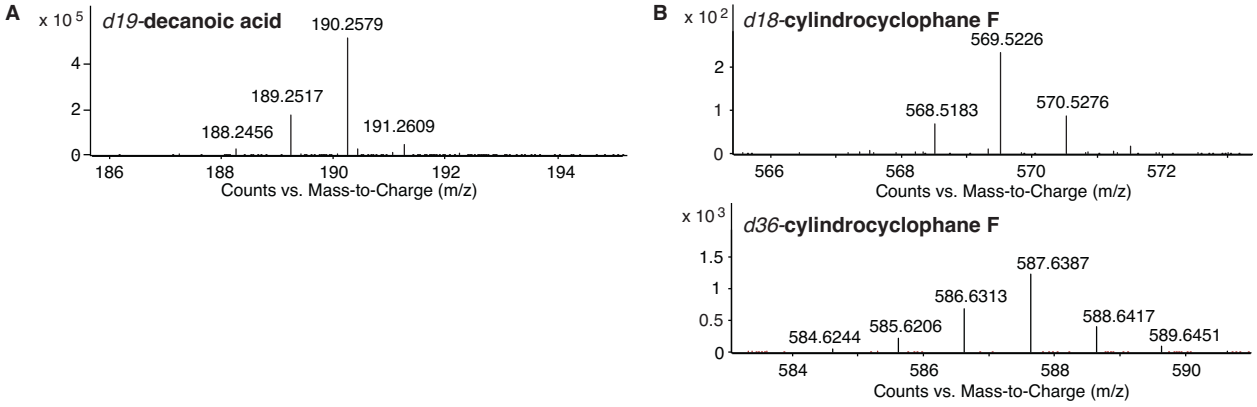


Figure 2.19: LC-MS m/z isotope distribution of deuterium labeled compounds in the feeding experiments. **A**) Isotope distribution of d_{19}-decanoic acid used for the feeding experiment. **B**) Isotope distribution of d_{18}- and d_{36}-cylindrocyclophane F detected in the feeding study samples.

Table 2.2: Isotope distribution of d_{19}-decanoic acid used for feeding experiments.

Compound formula	Ion	Expected Mass	Observed Mass	Error (ppm)	Peak Area	% of Mixture
$C_{10}H_3D_{17}O_2$	$[M-H]^-$	188.2458	188.2456	0	1617877	3.9%
$C_{10}H_2D_{18}O_2$	$[M-H]^-$	189.2520	189.2517	–1.59	10721466	25.8%
$C_{10}HD_{19}O_2$	$[M-H]^-$	190.2583	190.2579	–2.10	26411895	63.5%
$^{13}CC_9HD_{19}O_2$	$[M-H]^-$	191.2617	191.2614	–1.57	2874111	6.9%

Table 2.3: Isotope distribution of d_{18}-cylindrocyclophane F detected in the feeding experiments.

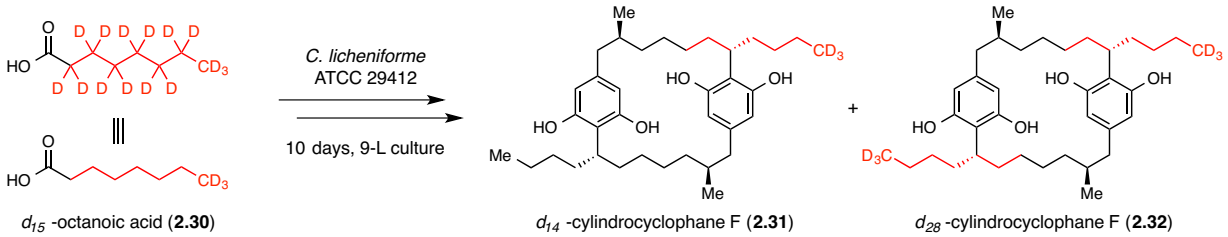
Compound formula	Ion	Expected Mass	Observed Mass	Error (ppm)	Peak Area	% of Mixture
$C_{36}H_{40}D_{16}O_2$	$[M-H]^-$	567.5110	–	–	Very low abundance	–
$C_{36}H_{39}D_{17}O_2$	$[M-H]^-$	568.5173	568.5183	1.76	1188	17.3%
$C_{36}H_{38}D_{18}O_2$	$[M-H]^-$	569.5236	569.5226	–1.76	4182	61.1%
$^{13}CC_{35}H_{38}D_{18}O_2$	$[M-H]^-$	570.5269	570.5276	1.23	1478	21.6%

Table 2.4: Isotope distribution of d_{36}-cylindrocyclophane F detected in the feeding experiments.

Compound formula	Ion	Expected Mass	Observed Mass	Error (ppm)	Peak Area	% of Mixture
$C_{36}H_{24}D_{32}O_2$	$[M-H]^-$	583.6114	583.6137	3.94	190	0.5%
$C_{36}H_{23}D_{33}O_2$	$[M-H]^-$	584.6177	584.6244	12.66	584	1.5%
$C_{36}H_{22}D_{34}O_2$	$[M-H]^-$	585.6240	585.6206	–5.81	2849	7.3%
$C_{36}H_{21}D_{35}O_2$	$[M-H]^-$	586.6303	586.6313	1.70	9546	24.5%
$C_{36}H_{20}D_{36}O_2$	$[M-H]^-$	587.6365	587.6387	3.74	18390	47.3%
$^{13}CC_{35}H_{20}D_{36}O_2$	$[M-H]^-$	588.6399	588.6417	3.06	5878	15.1%
$^{13}C_2C_{34}H_{20}D_{36}O_2$	$[M-H]^-$	589.6433	589.6451	3.05	1456	3.7%

2.2.7. Feeding studies with d_{14}-octanoic acid and d_{23}-dodecanoic acid

To gain more information about cylindrocyclophane biosynthesis, we performed additional feeding experiments in *C. licheniforme* with other potential precursors. We examined the possibility that the pathway might incorporate other free fatty acids through feeding studies using deuterium-labeled octanoic and dodecanoic acids. Feeding of d_{15}-octanoic acid (**2.30**) to *C. licheniforme* resulted in observation of masses corresponding to d_{14}- (**2.31**) and d_{28}-cylindrocyclophane F (**2.32**) by LC-MS analysis (**Scheme 2.14** and **Figure 2.20**).



Scheme 2.14: Production of d_{14}- and d_{28}-cylindrocyclophane F in the d_{15}-octanoic acid feeding experiment.

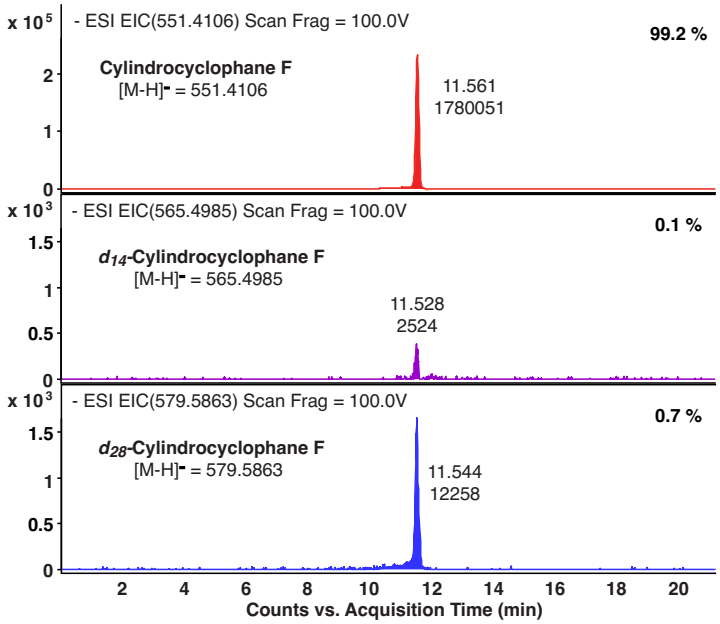
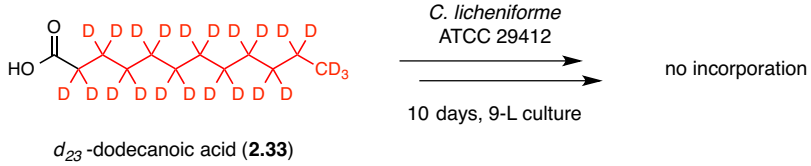


Figure 2.20: LC-HRMS extracted ion chromatograms of unlabeled and deuterium-labeled cylindrocyclophane F in the d_{15}-octanoic acid feeding experiment.

The result of the octanoic acid feeding experiment suggests that the fatty acid synthesis machinery first extends d_{15}-octanoic acid into d_{15}-decanoic acid, which can be used as a precursor for the cylindrocyclophanes. Feeding of d_{23}-dodecanoic acid (**2.33**) did not result in incorporation of the deuterium label into the cylindrocyclophane scaffold (**Scheme 2.15**). The lack of incorporation implies that d_{23}-dodecanoic acid is not catabolized into d_{19}-decanoic acid, which is consistent with the apparent lack of a fatty acid □-oxidation pathway in cyanobacteria.[30] In both of these feeding experiments, we did not detect any masses corresponding to deuterium-labeled cylindrocyclophane F analogs with shorter (**2.34**) or longer alkyl chains (**2.35**, see **Figure 2.21**), which are the expected products if d_{14}-octanoic acid (**2.30**) or d_{23}-dodecanoic acid (**2.31**) outcompetes decanoic acid and is used directly as a precursor for the cylindrocyclophanes. These results indicate that the fatty acid activation in cylindrocyclophane biosynthesis is very specific and free fatty acids with varying chain lengths cannot be used as precursors.



Scheme 2.15: Deuterium-labeled cylindrocyclophane F was not detected in d_{23}-dodecanoic acid feeding experiment.

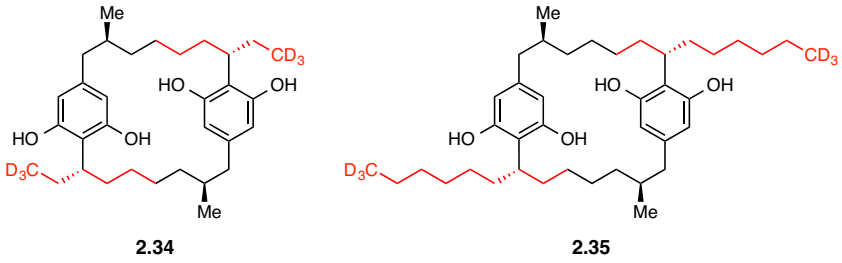
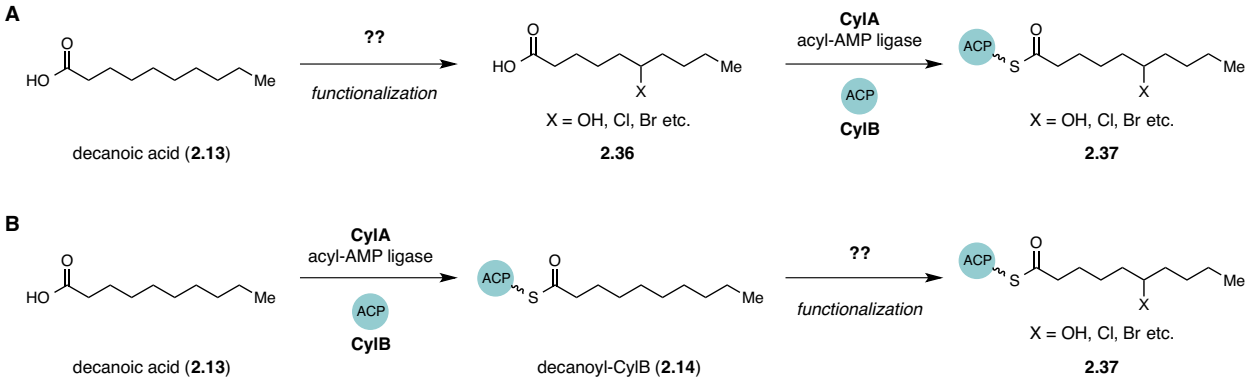


Figure 2.21: Possible structure of cylindrocyclophane F analogs with shorter or longer alkyl chains resulting from direct incorporation of d_{14}-octanoic acid or d_{23}-dodecanoic acid, respectively.

2.2.8. Feeding studies with possible biosynthetic intermediates

The d_{19}-decanoic acid feeding experiment showed that cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center at some point in the pathway. As discussed in more

detail in **Chapter 4**, one possible scenario for paracyclophane formation is that the functionalization of the eventual C7/C20 carbon centers occurs early in the biosynthesis, before (**Scheme 2.16A**) or after the fatty acid activation step (**Scheme 2.16B**). To assess this possibility, we performed a series of feeding studies using decanoyl-SNAC-derivatives (**2.38**-**2.43**, see **Figure 2.22**). These SNAC-derivatives (**2.38**-**2.43**), which mimic the structures of decanoyl-CylB (**2.14**) and potential functionalized decanoyl-CylB (**2.37**) biosynthetic intermediates, might be incorporated into the cylindrocyclophane scaffold. In addition, hydrolysis of the SNAC substrates should occur *in situ* to form the free fatty acids (**2.44**-**2.48**) during the course of the feeding experiments (**Scheme 2.17**). Thus, feeding experiments performed with the SNAC substrates would also allow us to assess the incorporation of the free decanoic acid derivatives simultaneously.



Scheme 2.16: Possible functionalization of the decanoic acid moiety early in the cylindrocyclophane biosynthesis. **A**) Functionalization of free decanoic acid prior to the fatty acid activation step. **B**) Functionalization of the decanoyl-CylB after the decanoic acid activation step.

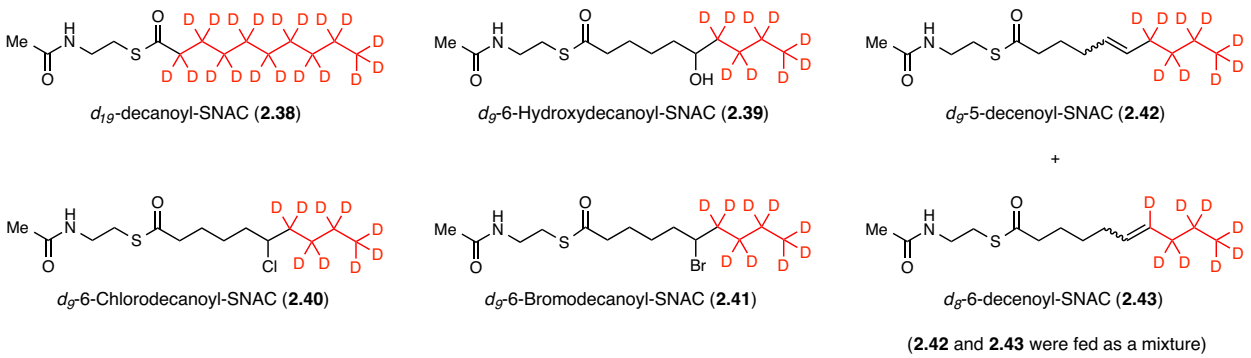
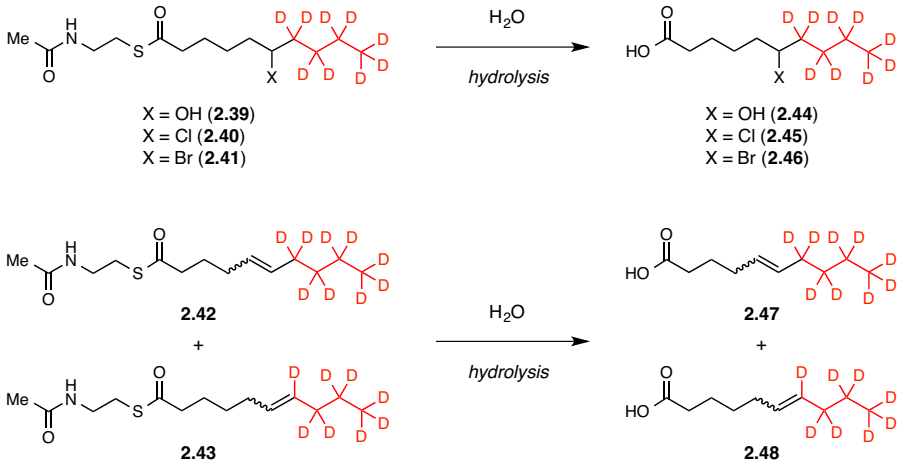


Figure 2.22: Structures of deuterium-labeled decanoyl-SNAC derivatives used for feeding studies in *C. licheniforme*.



Scheme 2.17: Hydrolysis of the SNAC substrates in the feeding experiments leads to *in situ* generation of the deuterium-labeled decanoic acid derivatives.

As we expected, feeding of d_{19}-decanoyl-SNAC (**2.38**) to the culture of *C. licheniforme* resulted in the production of d_{18}- (**2.28**) and d_{36}-cylindrocyclophane F (**2.29**). Feeding experiments performed with the functionalized decanoyl-SNAC substrates (**2.39-2.43**), however, did not lead to any incorporation of the deutereium label into the cylindrocyclophane scaffold. The lack of incorporation of the functionalized decanoyl-SNAC substrates (**2.39-2.43**) might suggest that the cylindrocyclophane assembly does not involve a pre-functionalization step early in the biosynthesis. We could not, however, exclude the possibility that cylindrocyclophane biosynthesis proceeds through an early pre-functionalization step solely based on the result of these feeding experiments. Although d_{19}-decanoyl-SNAC (**2.36**) was designed to mimic the structure of decanoyl-CylB (**2.14**), it was uncertain whether the downstream type I PKS CylD accepted the SNAC-derivative as a substrate. The incorporation of the deuterium label into the cylindrocyclophanes in the d_{19}-decanoyl-SNAC (**2.36**) feeding experiment might also have resulted from the hydrolysis of the SNAC substrate to form d_{19}-decanoic acid (**2.27**) *in vivo.* If the observed deuterium-labeling of the cylindrocyclophanes in feeding study with d_{19}-decanoyl-SNAC (**2.36**) is due to the incorporation of the free d_{19}-decanoic acid (**2.27**), then these feeding study results only shows that the pre-functionalized free decanoic acid derivatives cannot be activated by CylA for loading onto CylB. The

masses corresponding to free decanoic acid derivatives were detected by LC-MS in extracts from each of the feeding studies.

The other potential biosynthetic intermediate tested in a feeding study was d_3-resorcinol derivative (**2.49**, see **Figure 2.23A**), which is predicted to be the monomeric precursor to the cylindrocyclophanes. Addition of d_3-resorcinol to the culture of *C. licheniforme*, however, resulted in cytotoxicity and a dramatic color change of the culture from deep green to bright red (**Figure 2.23B**). Although we carried on the feeding experiment and analyzed the remaining cell extracts by LC-MS, we were unable to see incorporation of the deuterium label into the cylindrocyclophanes. This result could indicate that alkylated resorcinol **2.18** is not the true monomeric precursor to cylindrocyclophanes. However, we were uncertain whether the lack of incorporation of the deuterium label resulted from the disruption of proper metabolism in the producer due to cytotoxicity of d_3-resorcinol derivative **2.49**. The observed toxicity of the monomeric resorcinol would suggest that this potential precurse does not accumulate to high concentration under normal culturing conditions. In fact, alkyl resorcinol **2.18** does not appear to be present in the cell extract of the unfed cultures. This result suggests that if alkyl resorcinol **2.18** is truly a precursor to cylindrocyclophanes its formation is tightly coupled to the dimerization event to prevent accumulation of this cytotoxic intermediate.

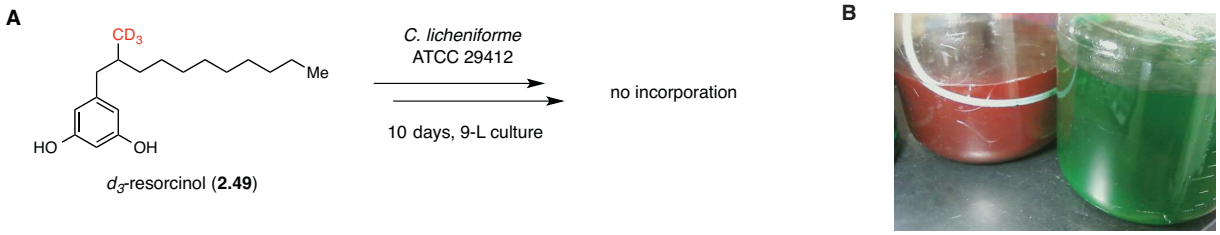
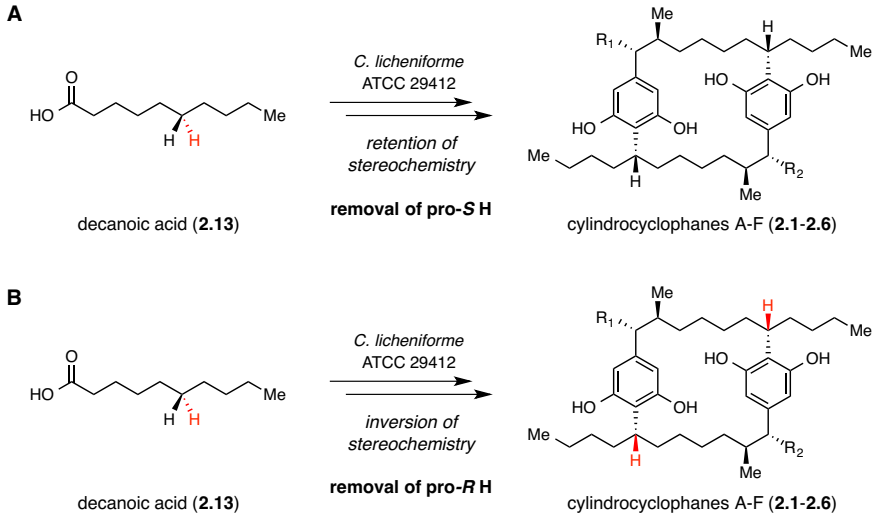


Figure 2.23: Feeding of d_3-resorcinol to *C. licheniforme.* **A**) Incorporation of d_3-resorcinol into the cylindrocyclophanes was not observed. **B**) Resorcinol appears to be toxic to the native producer, and a dramatic color change of the culture from green to bright red (culture on the left) is observed after several hours of feeding.

2.2.9. Feeding studies with stereoselectively labeled decanoic acids

One question we wished to answer regarding the cylindrocyclophane production was the stereoselectivity of the C–H activation and the C–C bond formation steps. From the result of the previous feeding experiments, we knew that the deuterium-labeled decanoic acid was incorporated into the final cylindrocyclophane scaffold. We next tested if it was possible to determine the stereoselectivity of the C–C bond formation using decanoic acid precursors that are stereoselectively deuterated at the C6 position where the eventual C–H activation and C–C bond formation occur.

Since the absolute stereochemistry of the cylindrocyclophanes was known,[9,31] we envisioned determining whether the biosynthesis of these molecules proceeded through retention or inversion of stereochemistry at the C7/C20 positions by observing stereoselective loss of either the pro-*S* or pro-*R* hydrogen atom in feeding experiments performed with stereoselectively deuterated precursors (**Scheme 2.18**). Similar feeding experiments have been used to determine the stereochemical outcome of the C–C bond formation in the streptorubin B biosynthesis, which proceeds with inversion of stereochemistry.[32] To examine the stereochemical outcome of paracyclophane formation in cylindrocyclophane biosynthesis, we synthesized the enantioselectively-deuterated decanoic acid substrates (**2.50** and **2.51**, see **Figure 2.24**).



Scheme 2.18: Investigation of the stereochemical outcome of paracyclophane formation in cylindrocyclophane biosynthesis. **A**) Cylindrocyclophane biosynthesis proceeds through retention of stereochemistry if the pro-*S* hydrogen is selectively removed. **B**) Cylindrocyclophane biosynthesis proceeds through inversion of stereochemistry if the pro-*R* hydrogen is selectively removed.

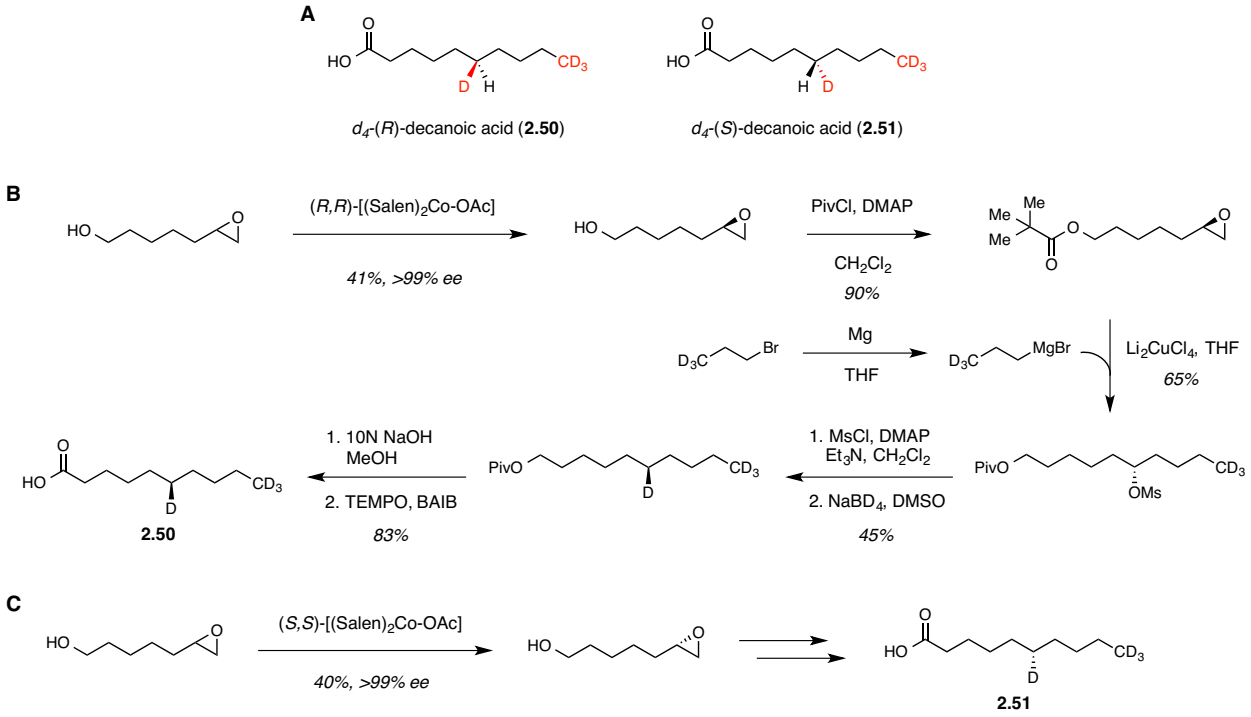
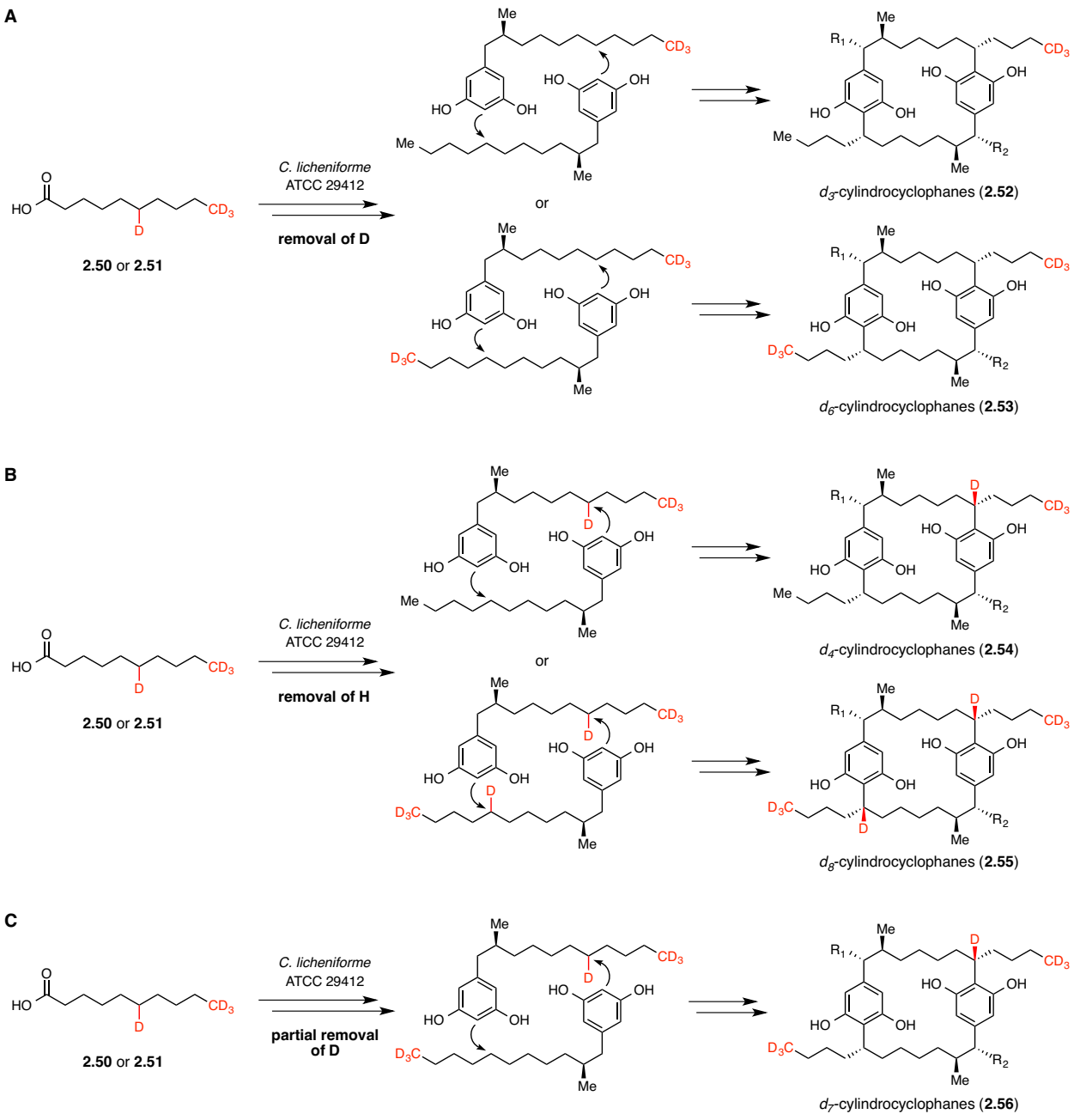


Figure 2.24: Synthesis of the stereoselectively deuterated decanoic acid substrates for the feeding experiments. **A**) Structures of the stereoselectively deuterated decanoic acid substrates. **B**) Synthetic route for the d_4-(*R*)-decanoic acid **2.50**. **C**) Synthesis for the d_4-(*S*)-decanoic acid **2.51** was performed following the same route as the d_4-(*R*)-decanoic acid, except starting with the epoxide of the other enantiomer.

In this feeding experiment, we expected to see five possible scenarios for the deuterium label incorporation. If the deuterium atom was removed during the C–H activation/C–C bond formation, then we would observe masses corresponding to d_3- (**2.52**) and d_6-cylindrocyclophanes (**2.53**, see **Scheme 2.19A**) from single or double incorporation, respectively. On the other hand, if the deuterium atom was not abstracted, then we would observe masses corresponding to d_4- (**2.54**) and d_8-cylindrocyclophanes (**2.55**, see **Scheme 2.19B**). Lastly, if the cylindrocyclophane biosynthetic pathway did not selectively remove hydrogen or deuterium atom at the eventual C7/C20 carbon center, then we might observe dimerization of one monomer with d_3-label and another monomer with d_4-label to form d_7-cylindrocyclophanes (**2.56,** see **Scheme 2.19C**).



Scheme 2.19: Possible outcomes of the stereoselectively labeled decanoic acid feeding studies. **A**) Selective removal of deuterium results in the production of d_3- and d_6-cylindrocyclophanes. **B**) Selective removal of hydrogen results in the production of d_4- and d_8-cylindrocyclophanes. **C**) Partial removal of deuterium may lead to coupling of d_3- and d_4-labeled precursors to form d_7-cylindrocyclophanes.

We observed incorporation of the deuterium labels into the cylindrocyclophane scaffold by LC-HRMS analysis of the culture extracts (**Figure 2.25A**). Interestingly, the level of incorporation appears to be much higher using a slightly modified protocol that was similar to the conditions used in the previous

feeding studies, except using smaller culture volume (1-L) and slightly higher substrate concentration (100 μM). Counter to our expectation, culture extracts fed with either d_4-(*R*)-decanoic acid (**2.50**) or d_4-(*S*)-decanoic acid (**2.51**) appear to have d_8-cylindrocyclophanes (**2.55**) and d_7-cylindrocyclophanes (**2.56**) as the major deuterium-labeled species (**Figure 2.25B**). This result indicates that for both d_4-(*R*)- and d_4-(*S*)-decanoic acid derivatives hydrogen atoms are removed preferentially and deuterium atoms are only occasionally removed. The trace amount of d_3- (**2.52**) and d_6-cylindrocyclophanes (**2.53**) seen likely resulted from the incomplete deuterium labeling of the decanoic acid substrates (**2.50** and **2.51**) used in the feeding experiments. The LC-MS analysis of **2.50** and **2.51** indicates that the deuterium-labeled substrates included approximately 4-5% of d_3-labeled species (**Figure 2.25C**).

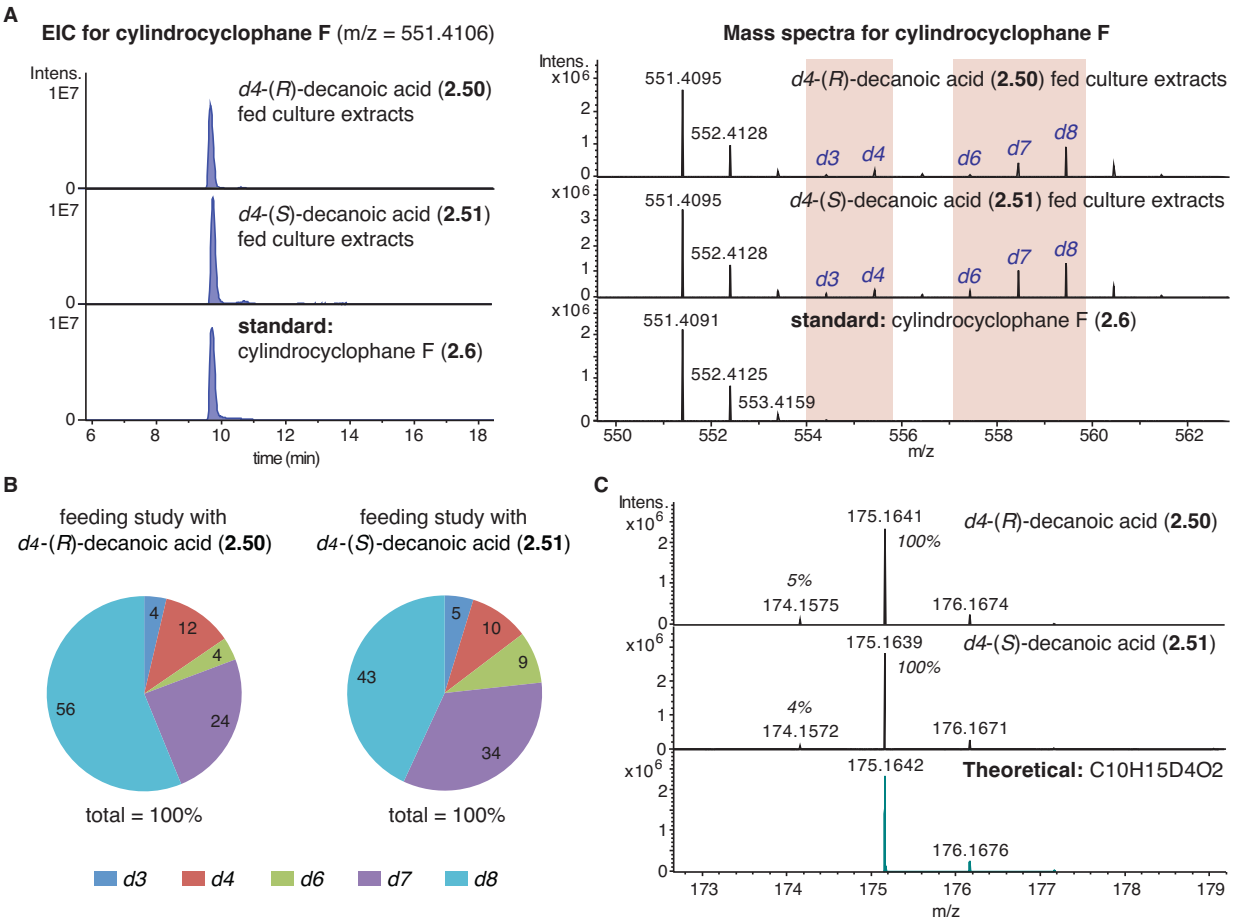
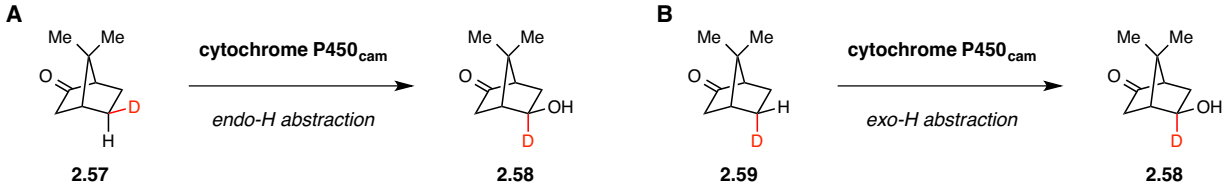


Figure 2.25: LC-MS analysis of the feeding studies performed with stereoselectively labeled decanoic acids. **A**) Extracted ion chromatograms and mass spectra of cylindrocyclophane F. Incorporation of deuterium label into cylindrocyclophane F is observed. **B**) Abundance (%) of d_3, d_4, d_6, d_7 and d_8-cylindrocyclophane F in the feeding studies. **C**) Mass spectra of deuterium labeled decanoic acids used for the feeding studies.

The preference for abstraction of hydrogen atoms instead of deuterium atoms observed for both substrates is likely a result of a strong deuterium kinetic isotope effect (KIE) associated with the C–H functionalization step. The similar level of hydrogen atom removal found with with either d_4-(*R*)- or d_4-(*S*)-decanoic acids (**2.50** and **2.51**) suggested that the enzyme(s) that catalyzes the C–H activation at the eventual C7/C20 carbon center is capable of abstracting both the pro-*S* and the pro-*R* hydrogen atoms. One example of an enzyme that displays similar behavior is the bacterial cytochrome P450 that catalyzes hydroxylation of camphor (**Scheme 2.20**).[33] Deuterium-labeling at either the 5-exo (**2.57**) or 5-endo (**2.59**) position of camphor results in an intramolecular isotope effect that leads to preferential abstraction of the hydrogen atom from either the endo or exo position.[33] Interestingly, hydroxylation only occurs at the 5-exo position (**2.58**) even when the 5-endo hydrogen atom is abstracted (**Scheme 2.20A**).[33] The C–H bond activating enzyme in the cylindrocyclophane biosynthesis may similarly catalyze functionalization of the C7/C20 carbon center stereoselectively despite the lack of preference for the stereochemistry of the initial hydrogen abstraction.



Scheme 2.20: Mechanistic studies of the cytochrome P450 camphor hydroxylase.[33] **A**) Deuterium labeling at the 5-exo position results in preferential abstraction of the 5-endo-hydrogen atom. The hydroxylation occurs at the 5-exo position, which results in the inversion of stereochemistry. **B**) Deuterium labeling at the 5-endo position results in preferential abstraction of the 5-exo-hydrogen atom and the stereochemistry is retained in the product.

While feeding of either d_4-(*R*)- or d_4-(*S*)-decanoic acids (**2.50** and **2.51**) led to similar distribution of d_3-, d_4-, d_6-, d_7- and d_8-cylindrocyclophanes (**2.52-2.56**), there appeared to be a slightly higher abundance of d_6- (**2.53**) and d_7-cylindrocyclophanes (**2.56**) in the culture extracts fed with d_4-(*S*)-decanoic acid. This observation could suggest the enzyme that catalyzes the C–H activation step prefers to abstract the pro-*S* hydrogen. If that is the case, then the cylindrocyclophane biosynthesis would proceed through retention or

double inversion of stereochemistry at the C7/C20 position. The result, however, is ambiguous and we cannot draw a definite conclusion on the stereochemical outcome of the cylindrocyclophane biosynthesis based on these feeding experiments. Stereoselectivity of the C–H activation and the C–C bond formation in cylindrocyclophane biosynthesis will be discussed in **Chapters 4** and **5**.

2.3. Conclusions

To study the cylindrocyclophane biosynthesis, we sequenced the genome of the producing organism, *C. licheniforme* ATCC 29412, and identified a candidate biosynthetic gene cluster through genome mining. The candidate *cyl* gene cluster was annotated based on bioinformatic analyses, and we formulated a biosynthetic hypothesis from the predicted enzyme functions. We then characterized the enzymes involved in the initiation and the termination of the type I PKS assembly line to test our biosynthetic hypothesis and confirm the connection between this gene cluster and cylindrocyclophane production.

Our biochemical characterization of CylA, CylB and CylI demonstrated that the construction of the resorcinol (**2.18**), the potential monomeric precursor of the cylindrocyclophanes, involves an intriguing combination of polyketide biosynthetic machinery. We found that cylindrocyclophane biosynthesis initiates with a size-selective recruitment of a fatty acid building block by CylA and CylB, followed by elaboration of the fatty acid by a type I modular PKS. A freestanding type III PKS CylI catalyzes the termination of the PKS assembly line by directly using an ACP-tethered nascent polyketide as a substrate.

Further verification of our biosynthetic hypothesis was provided by feeding experiments. Incorporation of d_{19}-decanoic acid (**2.27**) into the cylindrocyclophane scaffold *in vivo* confirmed that decanoic acid is the true precursor to the cylindrocyclophanes. This implies that the biosynthesis of the paracyclophanes core requires functionalization of an unactivated carbon center. In addition, the isotope distribution analysis of the labeled cylindrocyclophane F showed that only one deuterium is lost for incorporation of each d_{19}-decanoic acid unit. This suggests that the cylindrocyclophane biosynthesis does not involve an alkene or a ketone intermediate and likely requires a radical chemistry for the C–H activation step.

Additional feeding experiments provided some insight into the biosynthesis of the cylindrocyclophanes. The feeding study with d_{15}-octanoic acid (**2.30**) and d_{23}-dodecanoic acid (**2.33**) revealed that the cylindrocyclophane biosynthetic pathway is highly selective for fatty acid incorporation *in vivo*. We were unable to observe deuterium-label incorporation into the cylindrocyclophanes in feeding experiments performed with pre-functionalized decanoyl-SNAC derivatives (**2.39**-**2.43**). Consequently, it was unclear at this point whether cylindrocyclophane assembly involved pre-functionalization of the eventual C7/C20 carbon center early in the biosynthetic pathway. Significant cytotoxicity was observed in the feeding study using d_3-resorcinol (**2.49**), and the unfunctionalized resorcinol monomer (**2.18**) did not appear to accumulate to high concentration under normal culturing conditions. In addition, the feeding experiments performed with stereoselectively labeled d_4-(*R*)- or d_4-(*S*)-decanoic acids (**2.50** and **2.51**) revealed that the enzyme that performs the C–H activation step in cylindrocyclophane biosynthesis is able to abstract both the pro-*S* and the pro-*R* hydrogen atoms. The result of feeding studies with stereoselectively labeled d_4-(*R*) or d_4-(*S*)-decanoic acids (**2.50** and **2.51**) was difficult to interpret, since we can only observe the end product of multiple enzymatic steps in these experiments.

Through *in vitro* characterization of the biosynthetic enzymes and feeding studies, we have confirmed the involvement of the *cyl* gene cluster in cylindrocyclophane biosynthesis. Our discovery of the *cyl* gene cluster and verification of our biosynthetic hypothesis set the stage for future investigation of the cylindrocyclophane biosynthesis. Our next primary focus was to study to unusual interaction between the type I PKS CylH and type III PKS CylI and characterize the function of the C-terminal TE domain of CylH.

2.4. Materials and methods

2.4.1. Materials and general methods

Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). Recombinant plasmid DNA was purified with a Qiaprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean up were performed using an Illustra GFX PCR DNA and

Gel Band Purification kit from GE Healthcare. DNA sequencing was performed by Genewiz (Boston, MA). Nickel-nitrilotriacetic acid-agarose (Ni-NTA) resin was purchased from Qiagen. SDS-PAGE gels were purchased from BioRad. Protein concentrations were determined according to the method of Bradford using bovine serum albumin (BSA) as a standard.[1] Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm.

Analytical and preparative HPLC were performed on a Dionex Ultimate 3000 instrument (Thermo Scientific). Chiral HPLC analysis was performed in the Jacobsen lab in the Chemistry and Chemical Biology Department (CCB) at Harvard University using an Agilent analytical chromatograph with a commercial ChiralPak column. High-resolution mass spectral data for the synthetic compounds were obtained in the Magnetic Resonance Laboratory in CCB on Bruker MicroQTOF-QII fitted with a dual-spray electrospray ionization (ESI) source. The capillary voltage was set to 4.5 kV and the end plate offset to –500 V, the drying gas temperature was maintained at 190 ºC with a flow rate of 8 L/min and a nebulizer pressure of 21.8 psi. The liquid chromatography was performed using an Agilent Technologies 1100 series LC with 50% water and 50% acetonitrile as solvent. Isopropanol, methanol, and water used for LC-ESI-MS were B & J Brand High Purity Solvents (Honeywell Burdick & Jackson).

Proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded on an Agilent DD2-600 (600 MHz), Varian Inova-500 (500 MHz, 125 MHz), Varian-Mercury 400 (400 MHz, 100 MHz), or Varian-Mercury 300 (300 MHz) NMR spectrometers. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance as internal standard for ^{1}H ($CDCl_3$ = 7.26 ppm, CD_3OD = 3.31 ppm) and ^{13}C ($CDCl_3$ = 77.2 ppm, CD_3OD = 49.2 ppm). Data are reported as follows: chemical shift, multiplicity (s = singlet, bs = broad singlet, d = doublet, dd = doublet of doublets, t = triplet, m = multiplet), integration, and coupling constant. All chemicals were obtained from Sigma-Aldrich except for copper iodide (Strem Chemicals), sodium hydroxide (EMD Millipore), octanoic acid, decanoic acid, dodecanoic acid and tetradecanoic acid

(Alfa Aesar), malonyl-CoA (Santa Cruz Biotechnology), d_{19}-decanoic acid (Cambridge Isotope Laboratories), d_{15}-octanoic acid and d_{23}-dodecanoic acid (C/D/N Isotopes). Solvents were obtained from Sigma-Aldrich except hexanes (Macron Fine Chemicals), ethyl acetate and isopropanol (VWR), methanol and diethyl ether (EMD Millipore), and ethanol (KOPTEC). Solvents used for organic synthesis (THF, dichloromethane, and diethyl ether) were dried and degassed with the solvent columns. All NMR solvents were purchased from Cambridge Isotope Laboratories. *Cylindrospermum licheniforme* strain ATCC 29412 was purchased from the American Type Culture Collection.

2.4.2. Cultivation of *C. licheniforme* ATCC 29412 and isolation of cylindrocyclophane F

C. licheniforme ATCC 29412 was grown in 20 L polycarbonate carboy (Nalgene) containing 10 L of BG-11 medium (Sigma-Aldrich) at 25 ± 1 ºC with constant illumination (~1,000 lux) and aeration. After 20-25 days, the cyanobacterium was harvested by centrifugation (3,200 x*g* for 20 min at 4 ºC) and freeze-dried to yield an average of 0.14 g of biomass/L.

Cylindrocyclophane F was isolated using a procedure adapted from Moore and co-workers.[9] Freeze-dried *C. licheniforme* ATCC 29412 (1.36 g) was ground using a mortar and pestle, and the resulting paste was suspended in 7:3 ethanol/water (500 mL) and stirred overnight at 4 ºC. Insoluble cell debris was removed by centrifugation (3,200 x*g* for 10 min at 4 ºC) and the supernatant was concentrated *in vacuo*. The resulting crude extract was resuspended in methanol (100 mL), and the supernatant was decanted and concentrated *in vacuo.* The methanol extract was then purified by flash chromatography on silica gel, eluting with a gradient of dichloromethane, 19:1, 9:1, 4:1 and 1:1 mixtures of dichloromethane/acetone, and methanol. The 19:1, 9:1 and 4:1 dichloromethane/acetone fractions were combined and concentrated *in vacuo*. The residue was further purified by flash chromatography with hexanes followed by 19:1, 9:1 and 4:1 mixtures of hexanes/ethyl acetate to afford cylindrocyclophane F (4.6 mg, 7.6 μmoles). The characterization data for cylindrocyclophane F matched what has been reported previously.[9] TLC: R_f = 0.28 (silica gel, 4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{36}H_{55}O_4^-$ $[M–H]^-$, 551.4106; found 551.4113. ^{1}H NMR (500 MHz, CD_3OD) δ: 6.03 (s, 2H, aromatic C**H**), 5.99 (s, 2H, aromatic C**H**), 3.12 (m,

2H, benzylic C**H**), 2.60 (dd, 2H, *J* = 13.1, 3.7 Hz, benzylic C$\mathbf{H}_2$), 2.05-1.90 (m, 4H, C$\mathbf{H}_2$), 1.84 (t, 2H, *J* = 12.3 Hz, benzylic C$\mathbf{H}_2$), 1.65-1.55 (m, 2H, C**H**CH_3), 1.55-0.60 (m, 24H, C$\mathbf{H}_2$), 0.95 (d, 6H, *J* = 6.5 Hz, CHC$\mathbf{H}_3$), 0.83 (t, 6H, *J* = 7.2 Hz, C$\mathbf{H}_3$). ^{13}C NMR (125 MHz, CD_3OD) δ: 158.3, 157.3, 141.1, 116.3, 110.1, 108.2, 46.0, 37.0, 36.9 (2), 35.7, 35.1, 31.9, 30.8, 30.3, 24.1, 21.0, 14.7.

2.4.3. Isolation of genomic DNA from *C. licheniforme* and genome sequencing

Genomic DNA was isolated from ATCC 29412 using either the DNeasy Blood & Tissue Kit (Qiagen) following the protocol for gram-negative bacteria or the Ultra Clean Microbial DNA isolation kit from MoBio Laboratories (Carlsbad, CA). The shotgun sequencing and assembly were performed at GenoSeq (UCLA Genotyping and Sequencing Core) with the GS FLX Titanium system (Roche). The 454 sequencing reads were assembled into contigs with the GS De Novo Assembler software (Roche). Assembly resulted in 16.25 MB of non-redundant sequence distributed over 3225 contigs. The assembled data were converted into a local BLAST database for searches using stand-alone BLAST software (ver. 2.2.18) downloaded from NCBI. The putative cylindrocyclophane gene cluster was located on three contigs (1024, 111, and 112). The full-length sequence of the initial cylindrocyclophane gene cluster was obtained by PCR from *C. licheniforme* genomic DNA, cloning of PCR products, and sequencing. The sequences of contigs 1024, 111, and 112 were used to design specific PCR primers that would amplify regions of the genome spanning the three contigs or covering areas of the cluster suspected to contain sequencing errors (frameshifts). Blunt-ended PCR products (amplification with Pfu Turbo polymerase, Agilent) were ligated into the pCR4Blunt-TOPO vector using the Zero Blunt TOPO-PCR Cloning Kit (Invitrogen) and used to transform *E. coli* Top10 cells according to the manufacturers protocol. The sequences of the inserts enabled further assembly of contigs 1024, 111, and 112 into one continuous region of DNA.

The second genome sequencing and assembly was performed at Cofactor Genomics (St. Louis, MO). Short insert and long insert mate pair DNA libraries were sequenced using the HiSeq2000 (Illumina). Cluster generation and the subsequent sequencing were performed according to the cluster generation

manual and sequencing manual from Illumina (Cluster Station User Guide and Genome Analyzer Operations Guide). Base calls were generated using Casava 1.8.2 (Illumina), and the resulting demultiplexed sequence reads were filtered for low quality sequencing results. Assembly resulted in 16.6 MB of non-redundant sequence distributed over 200 scaffolds. The assembled data were converted into a local BLAST database for searches using Geneious Pro Version 7.1.6 (Biomatters). The cylindrocyclophane biosynthetic gene cluster was identified using the sequence of the initial *cyl* gene cluster as a query. The *cyl* gene cluster was found within scaffold 7 as one continuous sequence.

2.4.4. Annotation of the cylindrocyclophane biosynthetic gene cluster

Open reading frames (ORFs) in the *cyl* gene cluster were detected using FGENESB (Softberry), and the functions of each ORF was annotated based on BLAST[10] and conserved domain search.[11] We identified *cylA-L* in the original *cyl* gene cluster we obtained through 454 sequencing, and identified additional four ORFs (*cylM-P*) through the second sequencing attempt.

The domain annotations of the two PKS enzymes, CylD and CylH, were performed using the PKS/NRPS Analysis Website (http://nrps.igs.umaryland.edu/nrps/)[12] and the Conserved Domains Database at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The AT domains of CylH and CylD were analyzed by performing multiple sequence alignment with sequences of known AT domains from the avermectin biosynthetic pathway[13] using ClustalW in Geneious Pro. The MSA analysis showed that both the AT domains in CylD and CylH contain the conserved HAFH motif associated with malonyl-CoA incorporation.

2.4.5. Cloning of *cylA, cylB* and *cylI*

Table 2.5: Oligonucleotides used for cloning (restriction sites underlined).

Oligo	Sequence	Protein
cylA-for1	5' GAGATCCCATATGCATTTGCTACAACAAG 3'	N-His_6-CylA
cylA-rev-stop-2	5' GATCCTCGAGTCATTCTCTTTTTTTGGC 3'	
cylB-for-1	5' GAGATCCCATATGGAAATCGTAAATAGTC 3'	C-His_6-CylB
cylB-rev-nostop-1	5' GATCCTCGAGTGAATATTCTGCTTTAATTG 3'	
cylI-for-1	5' GAGATCCCATATGACTTATATAGTTTCTAC 3'	N-His_6-CylI
cylI-rev-stop-1	5' GATCCTCGAGTCACCATTTGAGTAAAATTGC 3'	

CylA, *cylB* and *cylI* were PCR amplified from *C. licheniforme* genomic DNA using the primers shown in **Table 2.5**. All PCR reactions (set up in triplicate) contained Phusion High-Fidelity PCR Master Mix (NEB), 2 ng DNA template, and 300 pmoles of each primer in a total volume of 30 uL. Thermocycling was carried out using the following parameters: denaturation for 1 min at 95 ºC, followed by 50 cycles of 30 sec at 95 ºC, 1 min at 50 ºC, 5 min at 70 ºC, and a final extension time of 10 min at 70 ºC. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled, and purified. Identical conditions were used for the amplification of CylA (forward primer **cylA-for-1** + reverse primer **cylA-rev-stop-2**), CylB (forward primer **cylB-for-1** + reverse primer **cylB-rev-nostop-1**) and CylI (forward primer **cylI-for-1** + reverse primer **cylI-rev-stop-1**) (all forward primers have *Nde*I restriction sites and all reverse primers have *Xho*I restriction sites).

Amplified fragments were digested with *Nde*I and *Xho*I (New England Biolabs) for 2.5 h at 37 ºC. Digests contained 2 µL water, 6 µL of NEB Buffer 4 (10x), 6 µL of bovine serum albumin (BSA, 10x), 40 µL of PCR product, 3 µL of *Nde*I (20 U/µL), and 3 µL of *Xho*I (20 U/µL). Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digests were ligated into linearized expression vector using T4 DNA ligase (New England Biolabs). *CylA* and *cylI* were ligated into the pET-28a vector to encode N-terminal His_6-tagged constructs,

and *cylB* was ligated into the pET-29b vector to encode a C-terminal His_6-tagged construct. Ligations were incubated at room temperature for 2 h and contained 3 μL water, 1 μL T4 Ligase buffer (10x), 1 μL digested vector, 3 μL digested insert DNA, and 2 μL T4 DNA Ligase (400 U/μL). A 5 μL aliquot of each ligation was used to transform a single tube of *E. coli* TOP10 cells (Invitrogen). The identities of the resulting pET-28a-*cylA*, pET-28a-*cylI*, and pET-29b-*cylB* constructs were confirmed by sequencing of purified plasmid DNA. These constructs were transformed into chemically competent *E.coli* BL21 (DE3) cells (Invitrogen) and stored at –80 °C as frozen LB/glycerol stocks.

2.4.6. Expression and purification of CylA, CylB and CylI

A 50 mL starter culture of pET-28a-*cylA*, pET-28a-*cylI* or pET-29b-*cylB* BL21 *E. coli* was inoculated from frozen stock and grown overnight at 37 °C in LB medium supplemented with 50 μg/ml kanamycin. Overnight cultures were diluted 1:50 into LB medium containing 50 μg/mL kanamycin (2 L for large-scale overexpressions). The cultures were incubated at 37 ºC with shaking for ~3.5 h, induced with 500 μM IPTG at OD_{600} = 0.5, and then incubated at 15 ºC for ~16 h.

Cells from 2 L of culture were pelleted by centrifugation (6,720 □*g* for 10 min) and resuspended in 40 mL of lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM $MgCl_2$, pH 8). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 10,000 psi, and the lysate was clarified by centrifugation (28,900 □*g* for 30 min). The supernatant was incubated with 2 mL of Ni-NTA resin and 5 mM imidazole for 2 h at 4 ºC. The mixture was centrifuged (3,200 rpm □*g* for 5 min) and the unbound fraction discarded. The Ni-NTA was resuspended in 10 mL of elution buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM $MgCl_2$, 5 mM imidazole, pH 8), loaded into a glass column, and washed with 10 mL of elution buffer. Protein was eluted from the column using a stepwise imidazole gradient in elution buffer (25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 200 mM) and collecting 2 mL fractions. SDS–PAGE analysis (4–15% Tris-HCl gel) was employed to ascertain the presence and purity of protein in each fraction. Fractions containing the desired protein were combined and dialyzed twice against 2 L of

storage buffer (20 mM Tris-HCl, 50 mM NaCl, 10% v/v glycerol, pH 8). Fractions containing CylB were combined and further purified by size-exclusion gel filtration on a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare) using an ÄKTA FPLC (AmershamPharmacia Biotech); fractions containing protein were combined and dialyzed against storage buffer (pH = 8) as described above. Solutions containing protein were frozen in liquid N_2 and stored at –80 ºC. This procedure afforded yields of 0.4 mg/L for N-His_6-CylA, 16 mg/L for C-His_6-CylB, and 20 mg/L for N-His_6-CylI.

2.4.7. BODIPY-CoA fluorescent phosphopantetheinylation assay

BODIPY-CoA[15] and Sfp[16] were prepared using previously reported procedures. Reaction mixtures (50 µL) contained 25 µM CylB, 150 µM Sfp, 25 µM BODIPY-CoA, 2 mM $MgCl_2$, and 50 mM Tris-HCl pH 8.0. Reactions were incubated in the dark at room temperature for 1 h. Reaction mixtures were diluted 1:1 in 2x Laemmli sample buffer (Bio-Rad), boiled for 10 min, and then separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on a 4-15% Tris-glycine gel. The gel was first imaged at λ=365 nm, then stained with Coomassie and imaged again.

2.4.8. HPLC assay for CylA fatty acid activation and loading onto CylB

Holo-CylB and decanoyl-CylB standards were prepared in separate 25 µL reactions. The reaction mixtures contained 50 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 100 µM CoA (for holo-CylB) or 1 mM decanoyl CoA (for decanoyl-CylB), 75 µM apo-CylB with 8 µM Sfp. Reactions were incubated at room temperature for 1 h.

A stock solution of *holo*-CylB was prepared prior to the assay by incubating 200 µM *apo*-CylB, 8 µM Sfp, 6 mM CoA, and 2 mM $MgCl_2$ in 50 mM Tris-HCl pH 8.0 for 1 h at room temperature. Loading assays contained 50 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 100 µM ATP, 50 µM decanoic acid (added as a 5 mM stock solution in DMSO), 50 µM *holo*-CylB and 100 nM CylA in a final assay volume of 100 µL. Samples were quenched with equal volume of acetonitrile and centrifuged (16,100 □*g* for 10 min at 4 ºC) before HPLC analysis on a Kinetex C8 column (2.6 µm, 100 Å, 100 x 3.0 mm, Phenomenex) at a flow

rate of 0.5 mL/min. The HPLC assay conditions were: 0 to 100% B in 30 min, hold at 100% B for 5 min, return to 0% B and hold for 8 min (solvent A = 0.1% trifluoroacetic acid in water, solvent B = 0.1% trifluoroacetic acid in acetonitrile).

2.4.9. Competition assay for evaluating size-selectivity of CylA/CylB fatty acid recruitment

A stock solution of *holo*-CylB was prepared prior to the assay by incubating 200 μM *apo*-CylB, 8 μM Sfp, 6 mM CoA, and 2 mM $MgCl_2$ in 50 mM Tris-HCl pH 8.0 for 1 h at room temperature. Competition assays contained 50 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 1 mM CoA, 1 mM ATP, 500 μM each of hexanoic, octanoic, decanoic, dodecanoic, and tetradecanoic acids, 50 μM *holo*-CylB and 5 μM CylA in a final assay volume of 50 μL. The assay was run alongside a negative control lacking CylA. Reaction mixtures were incubated at room temperature for 9.5 h, quenched with 100 μL of acetonitrile, and centrifuged (16,100 □*g* for 10 min). The quenched assay mixtures were also analyzed by LC-MS using a high accuracy Agilent ESI-TOF mass spectrometer in the Small Molecule Mass Spectrometry Facility, FAS Division of Science. A Varian PLRP-S column (2.0 mm ID, 1000 Å pore size, 10 cm length) was used to both desalt the sample and separate the unreacted fatty acids from the proteins. The LC assay conditions were: 5 to 90% B in 10 min (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). The averaged mass spectrum over the elution time contained several charge states for each of the protein species. This data was deconvoluted to neutral molecular weight using the maximum entropy algorithm (MaxEnt) that is integrated into the Agilent Bioconfirm software.

2.4.10. Biochemical characterization of CylI activity by HPLC assay

A typical assay contained 100 mM potassium phosphate buffer pH 8.0, 1 mM EDTA, 228 μM malonyl-CoA, 100 μM substrate □-ketoacyl-SNAC substrate (**2.26**) and 1 μM N-His_6-CylI in a final assay volume of 100 μL. An aqueous solution containing potassium phosphate buffer, EDTA, substrate and malonyl-CoA was prepared first. The assay mixture was vortexed, the enzyme was added, and the reaction was mixed gently with a pipet and incubated at room temperature. At various time points, 30 μL aliquots were removed from the reaction mixture, added to 60 μL of ice-cold acetonitrile, vortexed, incubated on ice for

10 min and centrifuged (16,100 □*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC (80 µL injection volume) on a Chromolith C18 RP-18 endcapped monolithic silica column (4.6 x 100 mm, EMD Millipore) at a flow rate of 1 mL/min. HPLC assay conditions were: 20% B in solvent A for 1 min, a gradient increasing to 100% B in solvent A over 14 min, 100% B for 10 min, a gradient decreasing to 20% B in solvent A over 2 min, 20% B for 5 min (solvent A = water, solvent B = acetonitrile).

2.4.11. Preparative scale CylI assay and isolation of resorcinol

To prepare sufficient quantities of the CylI product for characterization, 34 x 750 µL assays were set up. Each assay mixture contained 100 mM potassium phosphate buffer pH 8.0, 1 mM EDTA, 914 µM malonyl-CoA, 432 µM substrate **2.26** and 4.31 µM N-His_6-CylI. After 6 h at room temperate, each reaction was quenched with 750 µL of ice-cold acetonitrile and vortexed. All 34 reactions were combined into a separatory funnel and extracted with ethyl acetate (3 x 15 mL). The combined organic layers were washed with saturated aqueous solution of ammonium chloride (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude organic extract was purified by flash chromatography, eluting at 10–30% ethyl acetate in hexanes to afford resorcinol **2.18** (2.8 mg, 0.01 mmol, 91% yield). This material was determined to have 20% *ee* by chiral HPLC analysis (ChiralPak AD-H, 10% *i*PrOH in hexanes, 1 mL/min, 210 nm). ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 6.21 (d, 2H, J = 2.3 Hz, aromatic C**H**), 6.17 (t, 1H, J = 2.1 Hz, aromatic C**H**), 4.85 (bs, 2H, O**H**), 2.52 (dd, 1H, J = 13.3, 5.9 Hz, benzylic C$\mathbf{H}_2$), 2.22 (dd, 1H, J = 13.3, 8.7 Hz, benzylic C$\mathbf{H}_2$), 1.66 (m, 1H, C**H**), 1.39-1.20 (m, 15H, C$\mathbf{H}_2$), 1.12 (m, 1H, C$\mathbf{H}_2$), 0.88 (t, 3H, J = 6.9 Hz, C$\mathbf{H}_3$), 0.83 (d, 3H, J = 6.4 Hz, CHC$\mathbf{H}_3$). ^{13}C-NMR (125 MHz, $CDCl_3$) δ: 156.7, 145.2, 109.0, 100.4, 43.9, 37.0, 35.0, 32.2, 30.1, 30.0, 29.9, 29.6, 27.4, 23.0, 19.7, 14.4.

LC-MS analysis of the purified resorcinol was performed in the Saghatelian research labs in the Department of Chemistry and Chemical Biology, Harvard University on an Agilent G3250AA LC/MS TOF Mass Spectrometer fitted with a dual-spray electrospray ionization (ESI) source. The analysis was performed in negative ion mode. Liquid chromatography was performed using an Agilent Technologies

1200 series LC using a Phenomenex Gemini C18 reverse phase column (50 x 460 mm). LC-MS analysis using the following conditions: 100% solvent A for 2.1 min, a gradient increasing to 100% solvent B in solvent A over 7.9 min, 100% solvent B for 5 min, a gradient decreasing to 0% solvent B in solvent A over 0.1 min, 0% B for 5.9 min (solvent A = 95:5 water/methanol, solvent B = 80:15:5 isopropanol/methanol/water).

2.4.12. General procedure for feeding experiments and LC-MS analysis

Cylindrospermum licheniforme strain ATCC 29412 was grown in a 20-L carboy containing 9-L of BG-11 medium at 25 ± 1 ºC with constant illumination (~1,000 lux) and aeration. The substrate for feeding (**2.27**, **2.30**, **2.33**, **2.38-2.42** or **2.49**) was added as a 294 mM stock solution in DMSO 10 days after inoculation (65 μM final concentration). After 20–22 days of growth, the cells were harvested and the crude extract containing the cylindrocyclophanes was obtained using the procedure described in **2.4.2** for the isolation of cylindrocyclophane F. The resulting crude extract was dissolved in methanol (10 mL). The soluble material was decanted away from insoluble residues and used without further purification. High-resolution LC-MS analyses were performed using the same conditions described in **2.4.11**. 2 μL of the crude methanol extract was diluted into 198 μL of methanol for analysis (2 μL injection volume). All samples were centrifuged (16,100 □*g* for 10 min at 4 ºC) and the supernatants were removed for LC-MS analysis.

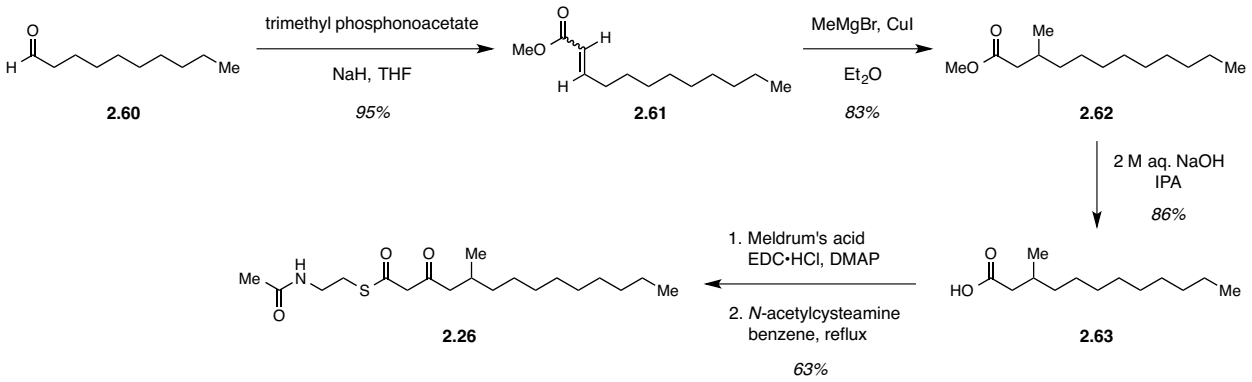
2.4.13. Feeding experiments with stereoselectively labeled decanoic acids

Cylindrospermum licheniforme strain ATCC 29412 was grown in a 1-L sterile plastic bottle containing 1-L of BG-11 medium at 25 ± 1 ºC with constant illumination (~1,000 lux) and aeration. The substrate for feeding (**2.50** or **2.51**) was added as a 500 mM stock solution in DMSO 10 days after inoculation (100 μM final concentration). After 20 days of growth, the cells were harvested and the crude extract containing the cylindrocyclophanes was obtained using the procedure described in **2.4.2** for the isolation of cylindrocyclophane F. The resulting crude extract was dissolved in methanol (2 mL). The soluble

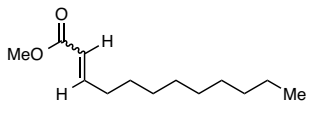
material was decanted away from insoluble residues and used without further purification. All samples were centrifuged (16,100 □*g* for 10 min at 4 ºC) and the supernatants were removed for LC-MS analysis.

LC-MS analysis was performed in the Small Molecule Mass Spectrometry Facility, FAS Division of Science. Mass spectral data were obtained on an Agilent 1290 Infinity UHPLC system (Agilent Technologies) and a maXis impact UHR time-of-flight mass spectrometer system (Bruker Daltonics Inc) equipped with an electrospray ionization (ESI) source. The analysis was performed in negative ion mode. Liquid chromatography was performed using a Phenomenex Gemini C18 reverse phase column (50 □ 460 mm) using the following conditions: 100% solvent A for 2.1 min, a gradient increasing to 100% solvent B in solvent A over 7.9 min, 100% solvent B for 5 min, a gradient decreasing to 0% solvent B in solvent A over 0.1 min, 0% B for 5.9 min (solvent A = 95:5 water/methanol, solvent B = 80:15:5 isopropanol/methanol/water).

2.4.14. Synthesis of the □-ketoacyl-SNAC substrate for CylI



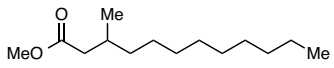
Scheme 2.21: Synthetic route for the □-ketoacyl-SNAC substrate **2.26**.



Methyl dodec-2-enoate (2.61): To an ice-cooled suspension of sodium hydride (0.525 g, 13.1 mmol, 60% dispersion in mineral oil) in THF (16.5 mL), trimethyl phosphonoacetate (1.94 ml, 11.9 mmol) was added dropwise over 5 min. The reaction mixture was warmed to room temperature and stirred for 1.5 h.

After re-cooling to 0 ºC, decanal (1.50 ml, 7.97 mmol) was added dropwise over 5 min to the reaction mixture, and the reaction mixture was allowed to gradually warm up to room temperature and stirred overnight. After re-cooling to 0 ºC, acetic acid (1.2 mL) was added and the reaction mixture was concentrated *in vacuo*. The crude reaction mixture was dissolved in ethyl acetate (75 mL) and washed with water (50 mL), saturated aqueous solution of sodium bicarbonate (50 mL) and brine (50 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo.* The crude product was purified by flash chromatography, eluting with 0–5% ethyl acetate in hexanes to afford methyl dodec-2-enoate **2.61** as a colorless oil (1.603 g, 7.55 mmol, 95%, *E*:*Z* 6:1). TLC: *Z* isomer R_f = 0.63 (silica gel, 19:1 hexanes/ethyl acetate); *E* isomer R_f = 0.48 (silica gel, 19:1 hexanes/ethyl acetate).

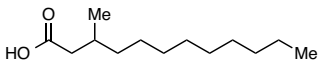
^{1}H and ^{13}C NMR data matched that reported previously.[34]



2.62

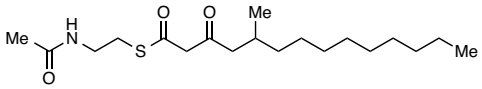
Methyl 3-methyldodecanoate (2.62): Methylmagnesium bromide (14.6 mL, 13.8mmol, 3.0 M solution in diethyl ether) was added dropwise over 5 min to a suspension of copper iodide (0.523 g, 2.75 mmol) in diethyl ether (10 mL) at –20 ºC, and the reaction mixture was stirred for 1.5 h. A solution of methyl dodec-2-enoate **2.61** (0.583 g, 2.75 mmol) in diethyl ether (3 mL) was added dropwise over 10 min and the reaction mixture was stirred at –20 ºC for 2.5 h. The reaction mixture was quenched by slow addition of methanol (10 mL) followed by saturated aqueous solution of sodium bicarbonate (40 mL). The aqueous layer was extracted with ethyl acetate (3 x 50 mL) and the combined organic layers were washed with brine (40 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 5 % ethyl acetate in hexanes to afford methyl 3-methyldodecanoate **2.62** as a colorless oil (0.409 g, 1.79 mmol, 83%). TLC: R_f = 0.37 (silica gel, 19:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_{29}O_2^+$ $[M+H]^+$, 229.2162; found, 229.2159. ^{1}H-NMR

(500 MHz, $CDCl_3$) δ: 3.65 (s, 3H, COOC**H**$_3$), 2.29 (dd, 1H, *J* = 14.6, 5.9, COC**H**$_2$), 2.09 (dd, 1H, *J* = 14.6, 7.8, COC**H**$_2$), 1.92 (m, 1H, C**H**), 1.14-1.35 (m, 16H, C**H**$_2$), 0.91 (d, 3H, *J* = 6.8, CHC**H**$_3$), 0.86 (t, 3H, *J* = 6.9, C**H**$_3$). ^{13}C-NMR (125 MHz, $CDCl_3$) δ: 174.0, 51.5, 41.9, 37.0, 32.1, 30.6, 30.0, 29.9, 29.8, 29.6, 27.1, 22.9, 20.0, 14.3.



2.63

3-Methyldodecanoic acid (2.63): To a solution of methyl 3-methyldodecanoate **2.62** (0.523 g, 2.290 mmol) in isopropanol (15 mL) was added 2 M aqueous sodium hydroxide (10 mL), and the resulting reaction mixture was stirred overnight at room temperature. The reaction mixture was acidified with 2 M aqueous hydrochloric acid (15 mL), and the aqueous layer was extracted with 3:1 chloroform/isopropanol (3 x 40 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to afford 3-methyldodecanoic acid **2.63** (0.421 g, 1.964 mmol, 86%) as a colorless oil. TLC: R_f = 0.65 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{13}H_{25}O_2^-$ $[M–H]^-$, 213.1860; found 213.1858. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 2.35 (dd, 1H, *J* = 14.6, 5.9 Hz, COC**H**$_2$), 2.14 (dd, 1H, *J* = 14.6, 7.8 Hz, COC**H**$_2$), 1.96 (m, 1H, C**H**), 1.35-1.16 (m, 16H, C**H**$_2$), 0.97 (d, 3H, *J* = 6.4 Hz, CHC**H**$_3$), 0.88 (t, 3H, *J* = 6.9 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz, $CDCl_3$) δ: 180.3, 41.9, 36.9, 32.2, 30.4, 30.0, 29.9, 29.8, 29.6, 27.1, 22.9, 19.9, 14.3.



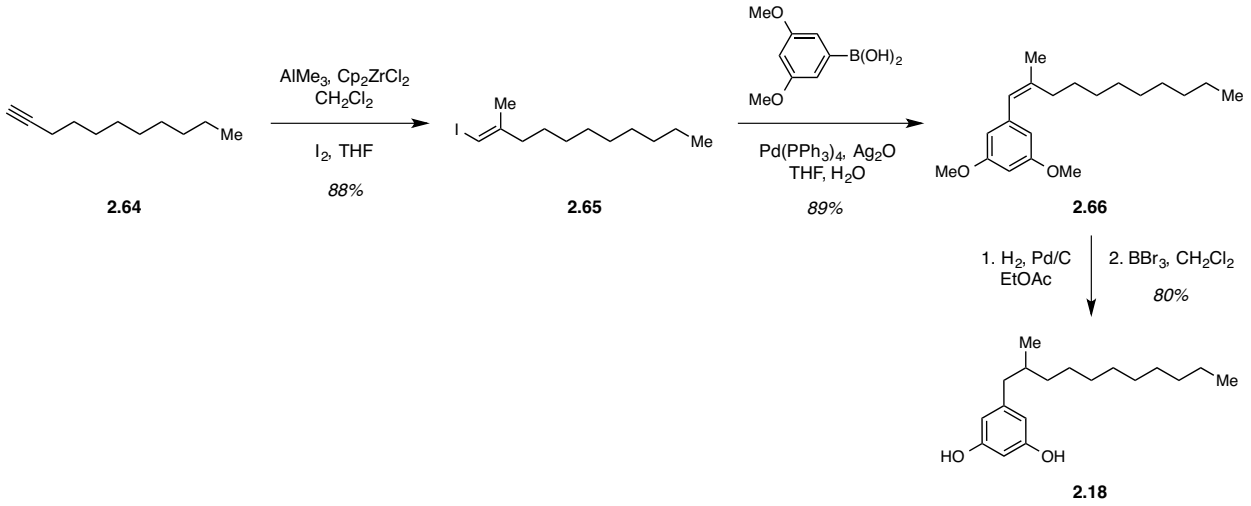
2.26

β-Ketoacyl-SNAC thioester (2.26): To solution of 3-methyldodecanoic acid **2.63** (100 mg, 0.467 mmol) in dichloromethane (6 mL) at 0 ºC was added 2,2-dimethyl-1,3-dioxane-4,6-dione (67.2 mg, 0.467 mmol) and 4-dimethylaminopyridine (171 mg, 1.40 mmol), followed by dropwise addition of a solution of EDC•HCl (97.0 mg, 0.509 mmol) in dichloromethane (3 mL) over 5 min. The reaction mixture was allowed to slowly warm to room temperature and stirred overnight. The reaction mixture was diluted with

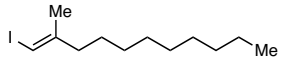
dichloromethane (30 mL) and the organic layer was washed with 1 M aq. HCl (3 x 30 mL) and brine (30 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to afford 2,2-dimethyl-5-(3-methyldodecanoyl)-1,3-dioxane-4,6-dione (156 mg, 92% crude recovery) as a yellow oil. The crude product was used for the next reaction without further purification. TLC: R_f = 0.35 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{19}H_{33}O_5^+$ $[M+H]^+$, 341.2323; found, 341.2331. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.06 (dd, 1H, *J* = 13.4, 6.1 Hz, C(OH)C**H**$_2$), 2.91 (dd, 1H, *J* = 13.2, 8.3 Hz, C(OH)C**H**$_2$), 2.03 (m, 1H, C**H**), 1.71 (s, 6H, C(C**H**$_3$)$_2$), 1.35 (m, 2H, C**H**$_2$), 1.30-1.15 (m, 14H, C**H**$_2$), 0.94 (d, 3H, *J* = 6.9 Hz, CHC**H**$_3$), 0.86 (t, 3H, *J* = 6.8 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz, $CDCl_3$) δ: 198.2, 104.9, 92.2, 42.7, 37.3, 32.2, 32.1, 30.0, 29.8, 29.8, 29.6, 27.1, 27.0, 22.9, 19.8, 14.3.

N-acetylcysteamine (103 mg, 0.860 mmol) was added to a solution of crude 2,2-4imethyl-5-(3-methyldodecanoyl)-1,3-dioxane-4,6-dione (146 mg, 0.430 mmol) in benzene (5 mL), and the reaction mixture was heated to reflux overnight. The reaction mixture was concentrated *in vacuo* and the crude product was purified by flash chromtography, eluting with 50–75% ethyl acetate in hexanes to afford β-ketoacyl-SNAC **2.26** as a colorless oil (97 mg, 0.271 mmol, 63% over two steps) as a mixture of tautomers. TLC: R_f = 0.33 (silica gel, 7.3 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{19}H_{36}NO_3S^+$ $[M+H]^+$, 358.2410; found, 358.2418. ^{1}H-NMR (500 MHz, $CDCl_3$) **enol form** δ: 12.59 (s, 1H, O**H**), 6.07 (s, 1H, N**H**), 5.43 (s, 1H, COC**H**CO), 3.46 (dt, 2H, *J* = 14.1, 6.4 Hz, NHC**H**$_2$), 3.07 (m, 2H, SC**H**$_2$),.18 (dd, 1H, *J* = 7.8, 2.3 Hz, COC**H**$_2$), 1.99 (m, 1H, COC**H**$_2$), 1.96 (s, 3H, COC**H**$_3$), 1.87 (m, 1H, C**H**), 1.39-1.12 (m, 16H, C**H**$_2$), 0.95-0.78 (m, 6H, C**H**$_3$); **keto form** δ: 6.07 (s, 1H, N**H**), 3.67 (s, 2H, SCOC**H**$_2$), 3.46 (dt, 2H, *J* = 14.1, 6.4 Hz, SNAC C**H**$_2$), 3.07 (m, 2H, SNAC C**H**$_2$), 2.49 (dd, 1H, *J* = 16.1, 5.4 Hz, COC**H**$_2$), 2.31 (dd, 1H, *J* = 16.6, 7.8 Hz, COC**H**$_2$), 1.96 (s, 3H, COC**H**$_3$), 1.87 (m, 1H, C**H**), 1.39-1.12 (m, 16H, C**H**$_2$), 0.95-0.78 (m, 6H, C**H**$_3$). ^{13}C-NMR (100 MHz, $CDCl_3$) δ: 202.4, 194.4, 192.6, 177.2, 170.6, 100.5, 57.8, 51.0, 42.8, 40.1, 39.4, 37.0, 37.0, 32.1, 31.3, 30.0, 29.9 (2), 29.8 (2), 29.5, 29.4, 29.3, 28.0, 27.1, 23.4, 23.4, 22.9, 19.9, 19.7, 14.3.

2.4.15. Synthesis of the resorcinol standard



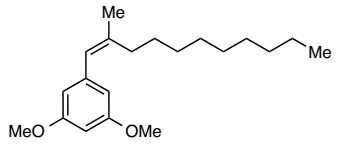
Scheme 2.21: Synthetic route for the resorcinol **2.18** standard.



2.65

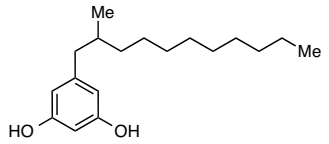
(*E*)-1-iodo-2-methylundec-1-ene (2.65): To a solution of $AlMe_3$ (0.99 mL, 2.0 M in hexanes, 1.98 mmol) and Cp_2ZrCl_2 (39 mg, 0.132 mmol) in dichloromethane (7 mL) at 0 ºC was added water (18 µL, 0.99 mmol). The reaction mixture was warmed to room temperature, stirred for 20 min, then cooled to 0 °C and a solution of undec-1-yne **2.64** (0.100 g, 0.66 mmol) in dichloromethane (1 mL) was added. The reaction mixture was stirred for 40 min at room temperature. A solution of I_2 (0.20 g, 0.79 mmol) in THF (2 mL) at 0 °C was then added and the reaction mixture was stirred for an additional 30 min at room temperature. The reaction mixture was poured into saturated aqueous solution of potassium carbonate (8 mL), filtered through a pad of Celite, and the filtrate was extracted with diethyl ether (2 x 10 mL). The combined organic layers were washed with brine (10 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography eluting with 10% diethyl ether in hexanes to afford (*E*)-1-iodo-2-methylundec-1-ene **2.65** (0.17 g, 0.58 mmol, 88%) as a colorless oil. TLC: R_f = 0.40 (silica gel, 9:1 hexanes/diethyl ether). HRMS (ESI): calc'd for $C_{12}H_{23}I^+$ $[M]^+$ 294.2155, found 294.2145. ^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.85 (s, 1H, C**H**I), 2.21-2.17 (m, 2H,

C**H₂**), 1.82 (s, 3H, C**H₃**), 1.43-1.40 (m, 2H, C**H₂**), 1.35-1.20 (m, 12H, C**H₂**), 0.88 (t, 3H, *J* = 7.1 Hz, $CH_2C\mathbf{H_3}$). ^{13}C NMR (125 MHz, $CDCl_3$) δ: 148.6, 74.5, 39.8, 32.1, 29.8, 29.6, 29.5, 29.3, 28.0, 24.1, 22.9, 14.4.



2.66

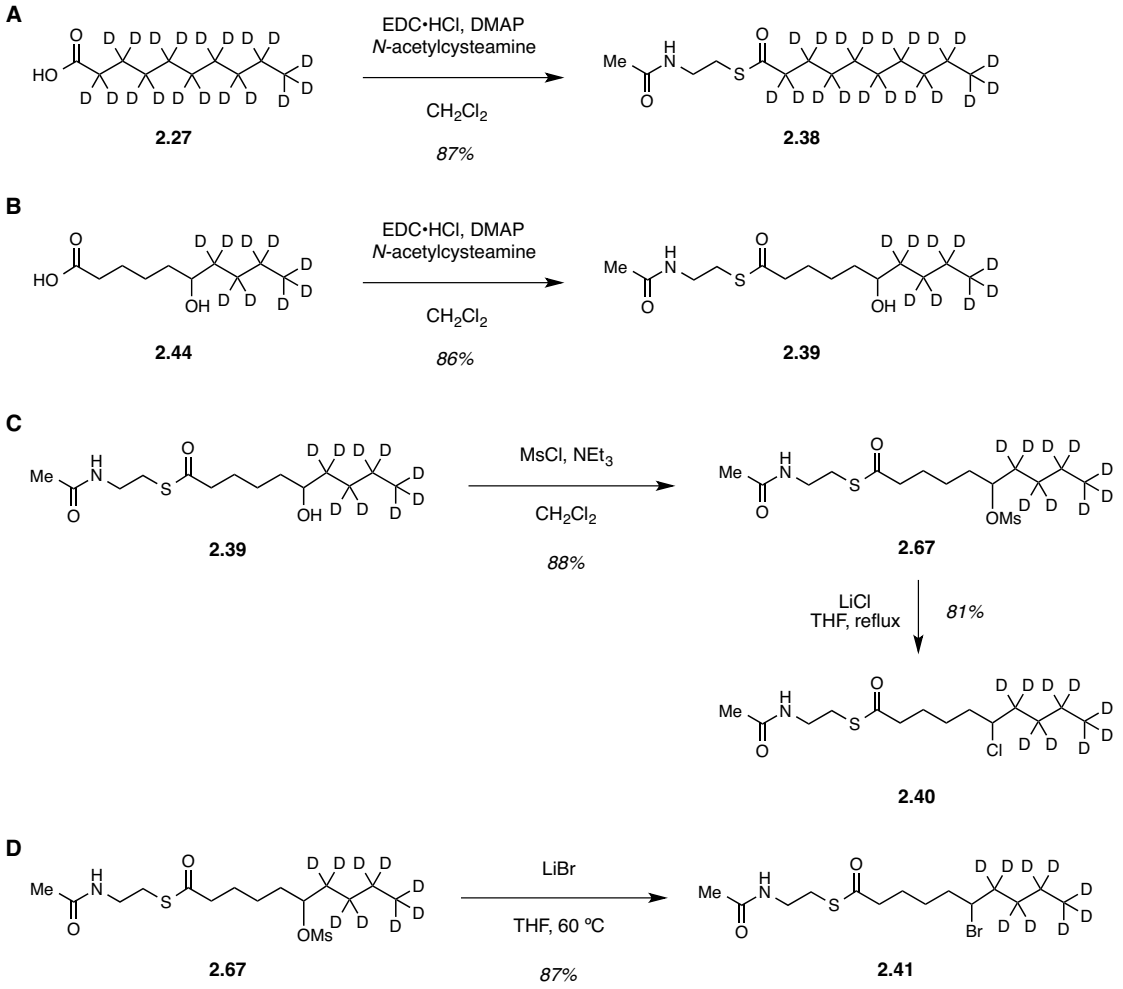
(*E*)-1,3-dimethoxy-5-(2-methylundec-1-en-1-yl)benzene (2.66): To a solution of (*E*)-1-iodo-2-methylundec-1-ene **2.65** (0.17 g, 0.58 mmol) in a mixture of THF (5 mL) and water (1 mL) were added 3,5-dimethoxyphenylboronic acid (0.12 g, 0.67 mmol), $Pd(PPh_3)_4$ (0.967 g, 0.058 mmol) and Ag_2O (0.034 g, 0.145 mmol). The reaction mixture was stirred at room temperature for 15 h before quenching with an addition of water (10 mL). The resulting solution was extracted with dichloromethane (2 x 15 mL). The combined organic layers were washed with brine (10 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography eluting with 20% diethyl ether in hexanes to afford (*E*)-1,3-dimethoxy-5-(2-methylundec-1-en-1-yl)benzene **2.66** (0.089 g, 89%) as a pale yellow oil. TLC: R_f = 0.50 (silica gel, 4:1 hexanes/diethyl ether). HRMS (ESI): calc'd for $C_{20}H_{33}O_2^+$ $[M+H]^+$, 305.2402; found, 305.2512. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 6.41 (s, 1H, aromatic C**H**), 6.40 (s, 1H, aromatic C**H**), 6.32 (t, 1H, *J* = 2.3 Hz, C**H**), 6.20 (s, 1H aromatic C**H**), 3.79 (s, 6H, C**H₃**), 2.14 (t, 2H, *J* = 7.3 Hz, CC**H₂**), 1.86 (s, 3H, C**H₃**), 1.51-1.48 (m, 2H, C**H₂**), 1.34-1.28 (m, 12H, C**H₂**), 0.89 (t, 3H, *J* = 7.1 Hz, $CH_2C\mathbf{H_3}$). ^{13}C-NMR (125 MHz, $CDCl_3$) δ: 160.7, 141.0, 140.3, 124.9, 107.2, 98.4, 55.5, 41.0, 32.2, 29.9, 29.8 (2C), 29.6, 28.2, 22.9, 18.2, 14.4.



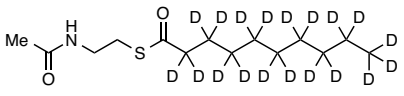
2.18

Resorcinol (2.18): To a solution of (*E*)-1,3-dimethoxy-5-(2-methylundec-1-en-1-yl)benzene **2.66** (0.14 g, 0.44 mmol) in dry ethyl acetate (6 mL) was added 10% palladium on carbon (0.095 g, 0.089 mmol). The reaction mixture was stirred at room temperature under an atmosphere of hydrogen gas (1 ATM) for 12 h. The crude mixture was then filtered through a pad of Celite and the filtrate was concentrated *in vacuo*. The crude residue was dissolved in dichloromethane (6 mL), cooled to 0 °C, and treated with BBr_3 (3.1 mL of a 1.0 M solution in dichloromethane, 3.1 mmol). The reaction mixture was stirred at room temperature for an additional 12 h, then cooled to 0 °C and quenched with the addition of saturated aqueous solution of potassium carbonate (2 mL). The resulting solution was diluted with water (5 mL) and extracted with dichloromethane (2 x 15 mL). The combined organic layers were washed with brine (10 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography eluting with 50% ethyl acetate in hexanes to give resorcinol **2.18** (0.099 g, 0.35 mmol, 80%) as a light yellow oil. TLC: R_f = 0.40 (silica gel, 5:5 hexanes/diethyl ether). HRMS (ESI): calc'd for $C_{18}H_{29}O_2^-$ $[M\text{-}H]^-$, 277.2246; found, 277.2176. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 6.27 (d, 2H, *J* = 2.3, aromatic C**H**), 6.22 (t, 1H, *J* = 2.1, aromatic C**H**), 6.08 (s, 2H, O**H**), 2.52 (dd, 1H, *J* = 13.3, 5.5, benzylic C$\mathbf{H}_2$), 2.18 (dd, 1H, *J* = 13.3, 8.7, benzylic C$\mathbf{H}_2$), 1.64 (m, 1H, C**H**), 1.38-1.21 (m, 15H, C$\mathbf{H}_2$), 1.14 (m, 1H, C$\mathbf{H}_2$), 0.91 (t, 3H, *J* = 6.9, $CH_2C\mathbf{H}_3$), 0.81 (d, 3H, *J* = 6.4, $CHC\mathbf{H}_3$).^{13}C-NMR (125 MHz, $CDCl_3$) δ: 156.2, 145.5, 109.4, 100.6, 43.8, 37.2, 34.9, 32.2, 30.2, 30.0, 29.9, 29.6, 27.4, 22.9, 19.5, 14.4.

2.4.16. Synthesis of pre-functionalized SNAC substrates for the feeding studies



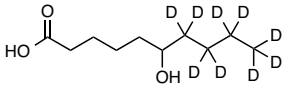
Scheme 2.23: Synthesis of deuterated SNAC substrates (**2.38-2.41**) for feeding experiments.



2.38

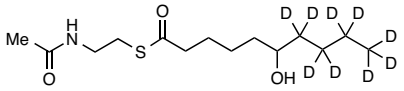
***D*$_{19}$-decanoyl-SNAC (2.38):** To d_{19}-Decanoic acid (100 mg, 0.52 mmol) was dissolved in dichloromethane (5 mL), EDC•HCl (220 mg, 1.2 mmol) was added. The reaction mixture was cooled to 0 ºC and stirred for 20 min. The flask was then charged with *N*-acetylcysteamine (74 µL, 0.70 mmol) and 4-dimethylaminopyridine (10 mg, 0.080 mmol), and the reaction mixture was warmed to room temperature. After 3 h, the reaction mixture was quenched with water (10 mL) and extracted with ethyl acetate (3 x 10

mL). The combined organic layers were washed with brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 75-100% ethyl acetate in hexanes to afford decanoyl-SNAC (130 mg, 0.45 mmol, 87%). TLC: R_f = 0.29 (1:3 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_9D_{19}NO_2S^+$ $[M+H]^+$, 293.3028; found 293.3029. ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 6.13 (s, 1H, N**H**), 3.38 (q, 2H, *J* = 6.4, NC**H**$_2$), 2.98 (t, 2H, *J* = 6.4, SC**H**$_2$), 1.93 (s, 3H, COC**H**$_3$). ^{13}C-NMR (125 MHz, CDCl3) δ: 200.0, 170.3, 39.5, 28.2, 22.9 (Note: The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets, which could not be identified.)



2.44

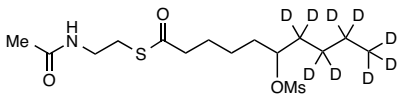
***D*$_9$-6-Hydroxydecanoic acid (2.44):** Synthesis was performed following previously reported procedures.[35,36] TLC: R_f = 0.35 (1:3 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{10}H_{10}D_9O_3^-$ $[M–H]^-$, 196.1905; found 196.1902. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 7.42-7.12 (bs, 1H, COO**H**), 3.51 (m, 1H, C**H**OH), 2.25 (t, 2H, *J* = 7.4 Hz, COC**H**$_2$), 1.65-1.43 (m, 2H, C**H**$_2$), 1.43-1.26 (m, 4H, C**H**$_2$). ^{13}C-NMR (125 MHz, $CDCl_3$) δ: 178.3, 71.5, 36.5 35.6 (m), 34.0, 26.4 (m), 25.0, 24.7, 21.3 (m), 12.8 (m).



2.39

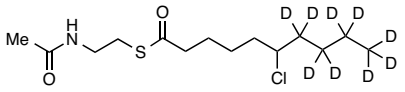
***D*$_9$-6-Hydroxydecanoyl-SNAC (2.39):** Synthesis was performed following the procedure described for the synthesis of *d*$_{19}$-decanoyl-SNAC **2.38**, except using *d*$_9$-6-hydroxydecanoic acid **2.44** (100 mg, 0.51 mmol) as the starting material to obtain *d*$_9$-6-hydroxydecanoyl-SNAC **2.39** (130 mg, 0.44 mmol, 86%). TLC: R_f = 0.50 (9:1 dichloromethane/methanol). HRMS (ESI): calc'd for $C_{14}H_{18}D_9NNaO_3S^+$ $[M+Na]^+$, 321.2169; found 321.2180. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 6.86 (s, 1H, N**H**), 3.41 (m, 1H, C**H**OH), 3.24 (q, 2H, *J* = 6.3 Hz, NC**H**$_2$), 2.87 (t, 2H, *J* = 6.6 Hz, S**H**$_2$), 2.43 (t, 2H, *J* = 7.4 Hz, SCOC**H**$_2$), 1.82 (s, 3H, COC**H**$_3$), 1.58-1.49 (m, 2H, C**H**$_2$), 1.36-1.21 (m, 4H, C**H**$_2$). ^{13}C-NMR (125 MHz; CDCl3): δ 199.6, 170.6,

70.9, 43.8, 39.2, 36.6, 36.0 (m), 28.2, 26.4 (m), 25.5, 24.8, 22.8, 21.2 (m), 12.7 (m). (Note: The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets.)



2.67

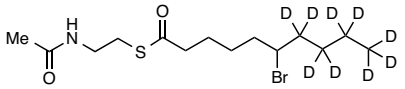
***D*$_9$-6-Mesyldecanoyl-SNAC (2.67):** *D*$_9$-6-hydroxydecanoyl-SNAC **2.39** (270 mg, 0.89 mmol) was dissolved in dichloromethane (30 mL) and the resulting solution was cooled to 0 ºC. Mesyl chloride (76 µL, 0.98 mmol) and triethylamine (0.15 mL, 1.1 mmol) were added to the reaction flask, and the reaction mixture was stirred at 0 ºC for 30 min. The reaction mixture was quenched with the addition of cold water (30 mL) and extracted with dichloromethane (3 x 30 mL). The combined organic layers were washed with brine (50 mL), dried over sodium sulfate, filtered and concentrated *in vacuo* to afford *d*$_9$-6-mesyldecanoyl-SNAC **2.67** (300 mg, 0.79 mmol, 88%). TLC: R_f = 0.46 (19:1 dichloromethane/ methanol). HRMS (ESI): calc'd for $C_{15}H_{21}D_9NO_5S_2^+$ $[M+H]^+$, 377.2125; found 377.2135. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 6.46 (s, 1H, N**H**), 4.57 (t, 1H, *J* = 6.0 Hz, C**H**OS), 3.29 (q, 2H, *J* = 6.3 Hz, NC**H**$_2$), 2.93-2.90 (t, 2H, *J* = 6.4 Hz, SC**H**$_2$), 2.92 (s, 3H, SC**H**$_3$), 2.48 (t, 2H, *J* = 7.4 Hz, SCOC**H**$_2$), 1.86 (s, 3H, COC**H**$_3$), 1.62-1.55 (m, 4H, C**H**$_2$), 1.38-1.28 (m, 2H, C**H**$_2$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 199.2, 170.5, 83.5, 43.5, 39.3, 38.5, 33.9, 33.0 (m), 28.4, 25.6 (m), 25.1, 24.0, 22.9, 21.0 (m), 12.6 (m). (Note: The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets.)



2.40

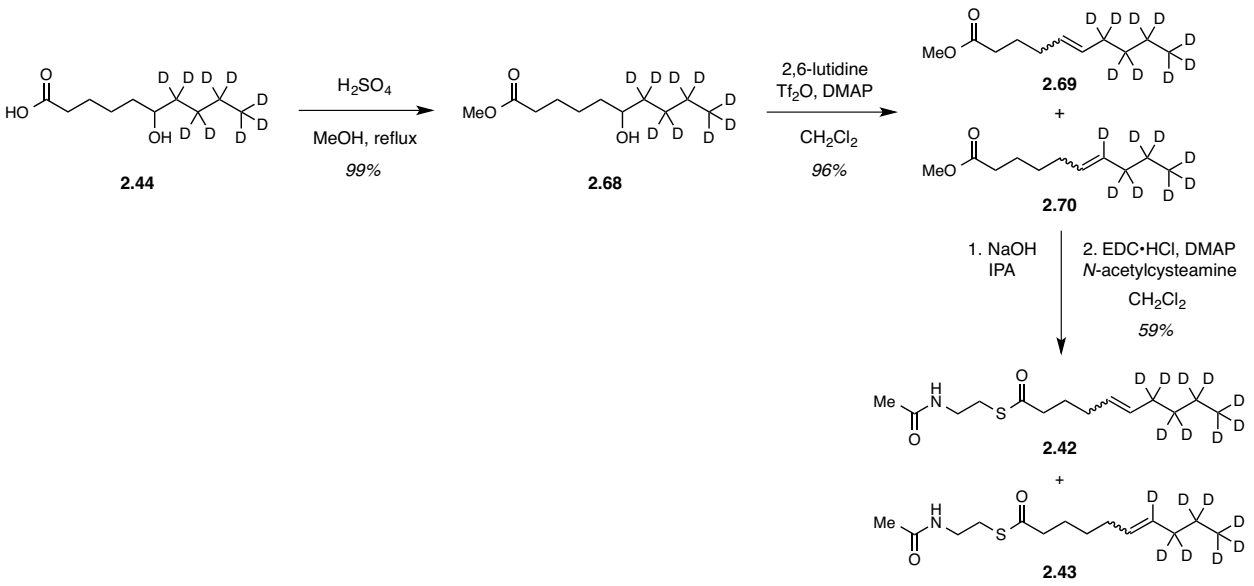
***D*$_9$-6-Chlorodecanoyl-SNAC (2.40):** *D*$_9$-6-mesyldecanoyl-SNAC **2.67** (80 mg, 0.22 mmol) was dissolved in THF (5 mL) and anhydrous lithium chloride (46 mg, 1.1 mmol) was added to the reaction flask. The reaction mixture was heated to reflux and held at this temperature for 8 h. The reaction mixture was cooled to room temperature, quenched with water (10 mL), and extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with brine (30 mL), dried over sodium sulfate and

concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 50-70% ethyl acetate in hexanes to afford d_9-6-chlorodecanoyl-SNAC **2.40** (54 mg, 0.18 mmol, 81%). TLC: R_f = 0.46 (3:7 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_{18}D_9ClNO_2S^+$ $[M+H]^+$, 317.2010; found 317.2011. ^{1}H-NMR (600 MHz; $CDCl_3$): δ 6.39 (s, 1H, N**H**), 3.78 (m, 1H, C**H**Cl), 3.33 (q, 2H, *J* = 6.2 Hz, NC**H**$_2$), 2.95 (t, 2H, *J* = 6.6 Hz, SC**H**$_2$), 2.51 (t, 2H, *J* = 7.4 Hz, SCOC**H**$_2$), 1.89 (s, 3H, COC**H**$_3$), 1.67-1.55 (m, 4H, C**H**$_2$), 1.50 (m, 1H, C**H**$_2$), 1.37 (m, 1H, C**H**$_2$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 199.5, 170.4, 63.6, 43.8, 39.5, 37.9, 37.2 (m), 28.4, 27.3 (m), 25.8, 25.0, 23.1, 20.9 (m), 12.8 (m). (Note: The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets.)

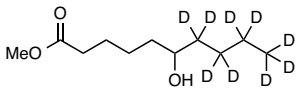


2.41

***D*$_9$-6-Bromodecanoyl-SNAC (2.41):** D_9-6-mesyldecanoyl-SNAC **2.67** (390 mg, 1.02 mmol) was dissolved in THF (20 mL) and lithium bromide (440 mg, 5.1 mmol) was added. The reaction mixture was heated to 60 ºC for 12 h. The reaction mixture was cooled to room temperature, quenched with water (10 mL), and extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with brine (30 mL), dried over sodium sulfate and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 50-70% ethyl acetate in hexanes to afford d_9-6-bromodecanoyl-SNAC **2.41** (320 mg, 0.89 mmol, 87%). TLC: R_f = 0.44 (3:7 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_{18}D_9BrNO_2S^+$ $[M+H]^+$, 361.1505; found 361.1523. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 6.55 (s, 1H, N**H**), 3.89 (t, 1H, *J* = 6.4 Hz, C**H**Br), 3.30 (q, 2H, *J* = 6.3 Hz, NC**H**$_2$), 2.92 (t, 2H, *J* = 6.6 Hz, SC**H**$_2$), 2.48 (t, 2H, *J* = 7.4 Hz, SCOC**H**$_2$), 1.87 (s, 3H, COC**H**$_3$), 1.73-1.68 (m, 2H, C**H**$_2$), 1.62-1.47 (m, 3H, C**H**$_2$), 1.36 (m, 1H, C**H**$_2$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 199.2, 170.4, 57.9, 43.7, 39.4, 38.4, 37.8, 28.3, 28.2 (m), 26.8, 24.8, 23.0, 20.7(m), 12.7 (m). (Note: The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets.)

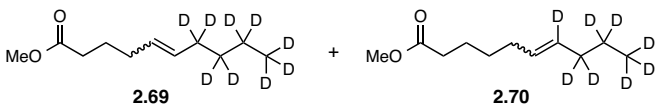


Scheme 2.24: Synthesis of the deuterated decenoyl-SNAC substrates (**2.41** and **2.42**) for feeding experiments.



2.68

***D_9*-Methyl-6-hydroxydecanoate (2.68):** *D_9*-6-hydroxydecanoic acid **2.44** (800 mg, 3.8 mmol) was dissolved in methanol (10 mL). Concentrated sulfuric acid (23 µL, 0.41 mmol) was added to the reaction flask, and the reaction mixture was warmed to reflux and held at this temperature for 1.5 h. The reaction mixture was then cooled to room temperature and concentrated *in vacuo*. The crude residue was purified by flash chromatography, eluting with 30-40% ethyl acetate in hexanes to afford *d_9*-methyl-6-hydroxydecanoate **2.68** as a colorless oil (790 mg, 3.8 mmol, 99%). TLC: R_f = 0.57 (1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{11}H_{14}D_9O_3^+$ $[M+H]^+$, 212.2207; found 212.2200. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.65 (s, 3H, OC**H**$_3$), 3.55 (m, 1H, C**H**OH), 2.31 (t, 2H, *J* = 7.5 Hz, COC**H**$_2$), 2.12 (s, 1H, O**H**), 1.67-1.57 (m, 2H, C**H**$_2$), 1.47-1.32 (m, 4H, C**H**$_2$). ^{13}C NMR (125 MHz; CDCl3): δ 174.3, 71.6, 51.6, 37.1, 34.1, 25.3, 25.0. (Note: ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets, which could not be identified.)

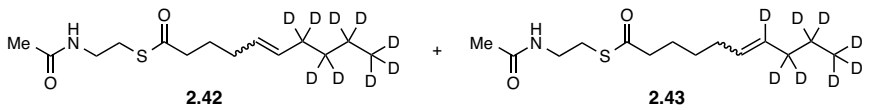


***D*$_9$-Methyl-5-decenoate (2.69)** and ***d*$_8$-methyl-6-decenoate (2.70):** *D*$_9$-methyl-6-hydroxydecanoate **2.68** (900 mg, 4.3 mmol) was dissolved in dichloromethane (20 mL) and the resulting solution was cooled to –15 ºC. 4-dimethylaminopyridine (52 mg, 0.43 mmol) and 2,6-lutidine (5 mL, 43 mmol) were added, followed by the dropwise addition of triflic anhydride (2.2 mL, 13 mmol) over 10 min. The reaction mixture was held at the same temperature for 30 min, then quenched with the addition of cold water (20 mL), and extracted with dichloromethane (3 x 20 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 5-10% diethyl ether in hexanes to afford *d*$_9$-methyl-5-decenoate **2.69** and *d*$_8$-methyl-6-decenoate **2.70** as a mixtures of isomers (5:1 mixture of *E*:*Z* isomers, 1:1 mixture of 5- and 6-decanoate, 790 mg, 4.1 mmol, 96%). TLC: R_f = 0.85 (1:1 hexanes/diethyl ether).

D$_9$-Methyl-5-decenoate **2.69**: ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.42-5.31 (m, 2H, C=C**H**), 3.66 (s, 3H, OC**H**$_3$), 2.29 (t, 2H, *J* = 7.6 Hz, COC**H**$_2$), 2.09-1.97 (m, 2H, C**H**$_2$), 1.71-1.59 (m, 2H, C**H**$_2$). ^{13}C-NMR (100 MHz; $CDCl_3$): ***E* alkene:** δ 174.3, 131.8-130.3, 129.9-128.4, 51.6, 34.1-33.5, 32.1-31.7, 25.0-24.6.; ***Z* alkene:** ^{13}C NMR (100 MHz; $CDCl_3$): δ 174.3, 131.8-130.3, 129.9-128.4, 51.6, 34.1-33.5, 25.0-24.6, 26.6-26.9.

D$_8$-Methyl-6-decenoate **2.70**: ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.42-5.31 (m, 1H, C=C**H**), 3.66 (s, 3H, OC**H**$_3$), 2.29 (t, 2H, *J* = 7.6 Hz, COC**H**$_2$), 2.09-1.97 (m, 2H, C**H**$_2$), 1.71-1.59 (m, 2H, C**H**$_2$), 1.40-1.34 (m, 2H, C**H**$_2$). ^{13}C-NMR (100 MHz; $CDCl_3$): ***E* alkene:** δ 174.3, 129.9-128.4, 51.6, 34.1-33.5, 32.1-31.7, 29.2, 25.0-24.6. ***Z* alkene:** ^{13}C-NMR (100 MHz; $CDCl_3$): δ 174.3, 129.9-128.4, 51.6, 34.1-33.5, 29.3, 26.6-26.9, 25.0-24.6.

(Note: the ^{13}C peaks with ranges contain multiple peaks (2-4) of the possible *E/Z* isomers of 5- and 6-decanoyl methyl esters. The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets, which could not be identified.)



***D*$_9$-5-Decenoyl-SNAC (2.42)** and ***d*$_8$-6-decenoyl-SNAC (2.43):** A mixture of *d*$_9$-methyl-5-decenoate **2.69** and *d*$_8$-methyl-6-decenoate **2.70** (790 mg, 4.1 mmol) was dissolved in isopropanol (20 mL) and 2 M aqueous sodium hydroxide (20 mL) was added. The reaction mixture was stirred overnight at room temperature. The reaction mixture was then acidified with 2 M aqueous hydrochloric acid (30 mL) and the product was extracted with chloroform (3 x 50 mL). The combined organic layers were dried over sodium sulfate, filtered, and concentrated *in vacuo*. The crude decenoic acid was used directly for the next reaction without further purification. TLC: R_f = 0.48 (1:1 hexanes/diethyl ether). Decenoic acid (4.1 mmol) was dissolved in dichloromethane (18 mL) and the solution was cooled to 0 ºC. EDC•HCl (2.4 g, 12.6 mmol) was added and the reaction mixture was reacted at 0 ºC for 20 min. Then, 4-dimethylaminopyridine (76 mg, 0.62 mmol) and *N*-acetylcysteamine (840 µL, 7.9 mmol) were added and the reaction mixture was warmed to room temperature and stirred for an additional 3 more hours before quenching with water (20 mL) and extracted with dichloromethane (3 x 30 mL). The combined organic layers were dried over sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 40-60% ethyl acetate in hexanes to afford *d*$_9$-5-decenoyl-SNAC **2.42** and *d*$_8$-6-decenoyl-SNAC **2.43** as a mixture of isomers (5:1 mixture of *E*:*Z* isomers, 1:1 mixture of 5- and 6-decanoate, 680 mg, 2.44 mmol, 59%). TLC: R_f = 0.21 (2:3 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_{17}D_9NO_2S^+$ $[M+H]^+$, 281.2244; found 281.2253.

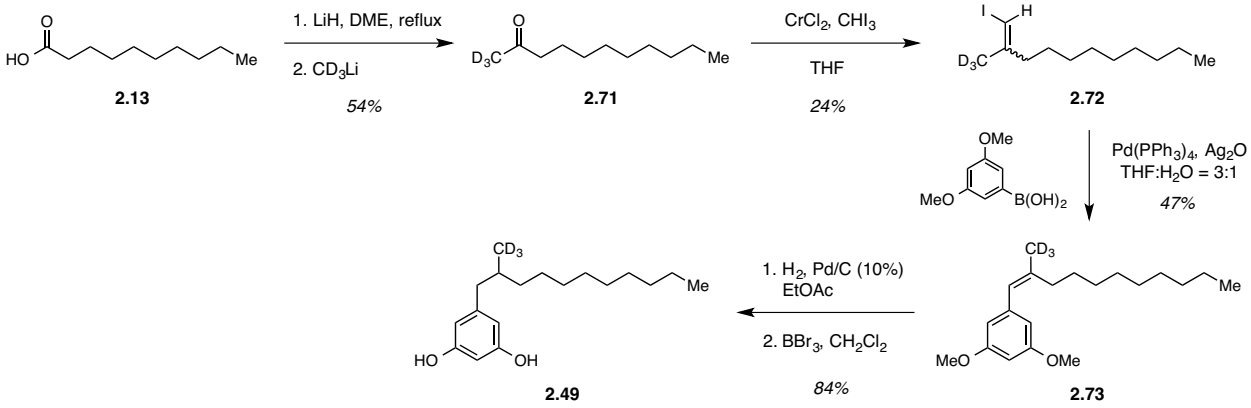
D$_9$-5-Decenoyl-SNAC **2.42**: ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.98 (1, 1H, N**H**), 5.42-5.29 (m, 2H, C=C**H**), 3.41 (q, 2H, *J* = 6.2 Hz, NC**H**$_2$), 3.00 (t, 2H, *J* = 6.5 Hz, SC**H**$_2$), 2.54 (t, 2H, *J* = 7.5 Hz, COC**H**$_2$), 2.07-1.94 (m, 2H, C**H**$_2$), 1.94 (s, 3H, COC**H**$_3$), 1.73-1.63 (m, 2H, C**H**$_2$). ^{13}C-NMR (125 MHz; $CDCl_3$): ***E***

alkene: δ 200.2, 170.4, 132.1-131.5, 129.5-128.1, 44.1-43.5, 39.8, 32.2-31.8, 28.5, 25.7-25.2, 23.3. ***Z*** **alkene:** δ 200.2, 170.4, 132.1-131.5, 129.5-128.1, 44.1-43.5, 39.8, 28.5, 26.8-26.4, 25.7-25.2, 23.3.

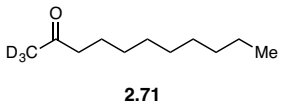
D_8-6-Decenoyl-SNAC **2.43:** ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.98 (1, 1H, N**H**), 5.42-5.29 (m, 1H, C=C**H**), 3.41 (q, 2H, *J* = 6.2 Hz, NC**H**$_2$), 3.00 (t, 2H, *J* = 6.5 Hz, SC**H**$_2$), 2.54 (t, 2H, *J* = 7.5 Hz, COC**H**$_2$), 2.07-1.94 (m, 2H, C**H**$_2$), 1.94 (s, 3H, COC**H**$_3$), 1.73-1.63 (m, 2H, C**H**$_2$), 1.39-1.35 (m, 2H, C**H**$_2$). ^{13}C-NMR (125 MHz; $CDCl_3$): ***E*** **alkene:** δ 200.2, 170.4, 129.5-128.1, 44.1-43.5, 39.8, 32.2-31.9, 29.1, 28.5, 25.7-25.2, 23.3. ***Z*** **alkene:** δ 200.2, 170.4, 129.5-128.1, 44.1-43.5, 39.8, 28.9, 28.5, 26.8-26.4, 25.7-25.2, 23.3.

(Note: the exact ^{13}C peaks with ranges contain multiple peaks of the possible *E/Z* isomers of 5- and 6-decanoyl methyl esters. The ^{13}C peaks of the carbon atoms bonded to the deuterium atoms are broad multiplets, which could not be identified.)

2.4.17. Synthesis of *d_3*-resorcinol



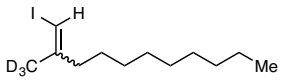
Scheme 2.25: Synthesis of *d_3*-resorcinol **2.49** for the feeding experiment.



2.71

***D_3*-Undecan-2-one (2.71):** To a suspension of LiH (73 mg, 8.7 mmol) in dimethoxyethane (20 mL), decanoic acid (1 g, 5.8 mmol) was added as a solution in dimethoxyethane (50 mL). The reaction mixture

was refluxed for 2.5 h, and then cooled to room temperature. To the reaction mixture, d_3-methyllithium (0.5 M in diethyl ether, 17.4 mL, 8.7 mmol) was added dropwise over 10 min, and the reaction mixture was stirred overnight. The reaction mixture was quenched with 1 M HCl solution (100 mL) and extracted with diethyl ether (3 × 100 mL). The combined organic layers were washed with saturated aqueous solution of sodium thiosulfate (100 mL) and brine (100 mL), dried over anhydrous sodium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 5-10% diethyl ether in hexanes to afford d_3-undecan-2-one **2.71** a colorless oil (540 mg, 3.1 mmol, 54% yield). TLC: R_f = 0.28 (9:1 hexanes/diethyl ether). ^{1}H NMR (300 MHz, $CDCl_3$) δ: 2.34 (t, 2H, *J* = 7.6, COC**H**$_2$), 1.49 (m, 2H, C**H**$_2$), 1.12-1.29 (m, 12H, C**H**$_2$), 0.80 (t, 0.80, *J* = 6.6, CH_2C**H**$_3$). ^{13}C NMR (125 MHz, CDCl3) δ: 209.1, 59.2 (m), 43.6, 31.7, 29.3, 29.3, 29.1, 29.0, 23.7, 22.5, 13.9. (Note: ^{13}C peak of the carbon atom bonded to the deuterium atoms is broad multiplet, which is difficult to identify).



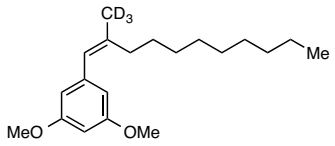
2.72

***D*$_3$-1-iodo-2-methylundec-1-ene (2.72):** d_3-undecan-2-one **2.71** (535 mg, 3.09 mmol) dissolved in THF (20 mL), $CrCl_2$ (1.51 g, 12.3 mmol) was added. To the reaction mixture, a solution of iodoform (2.43 g, 6.17 mmol) in THF (10 mL) was added dropwise over 10 min, and the reaction mixture was stirred overnight. The reaction mixture was concentrated *in vacuo* and the crude product was purified by flash chromatography, eluting in 100% hexanes to afford d_3-1-iodo-2-methylundec-1-ene **2.72** as a mixture of *E* and *Z* isomers (220 mg, 0.74 mmol, 24 % yield). The starting material was also recovered (278 mg, 1.61 mmol, 52%). TLC: R_f = 0.88 (9:1 hexanes/diethyl ether).

***E* alkene:** ^{1}H NMR (300 MHz, $CDCl_3$) δ: 5.85 (s, 1H, alkene C**H**I), 2.18 (t, 2H, *J* = 7.2, C**H**$_2$) 1.41 (m, 2H, C**H**$_2$), 1.15-1.35 (m, 12H, C**H**$_2$), 0.88 (t, 3H, *J* = 6.0, CH_2C**H**$_3$). ^{13}C NMR (125 MHz, CDCl3) δ: 148.2, 74.3, 39.6, 31.9, 29.5, 29.4, 29.1, 27.7, 27.0, 22.7, 14.1.

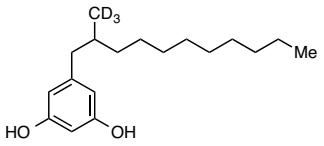
***Z* alkene:** ^{1}H NMR (300 MHz, $CDCl_3$) δ: 5.81 (s, 1H, alkene C**H**I), 2.18 (t, 2H, *J* = 7.2, C$\mathbf{H}_2$) 1.41 (m, 2H, C$\mathbf{H}_2$), 1.15-1.35 (m, 12H, C$\mathbf{H}_2$), 0.88 (t, 3H, *J* = 6.0, $CH_2C\mathbf{H}_3$). ^{13}C NMR (125 MHz, CDCl3) δ: 147.8, 73.8, 38.6, 31.9, 29.5, 29.4, 29.1, 27.7, 27.0, 22.7, 14.1.

(Note: ^{13}C peaks of the carbon atom bonded to the deuterium atoms is broad multiplet, which could not be identified).



2.73

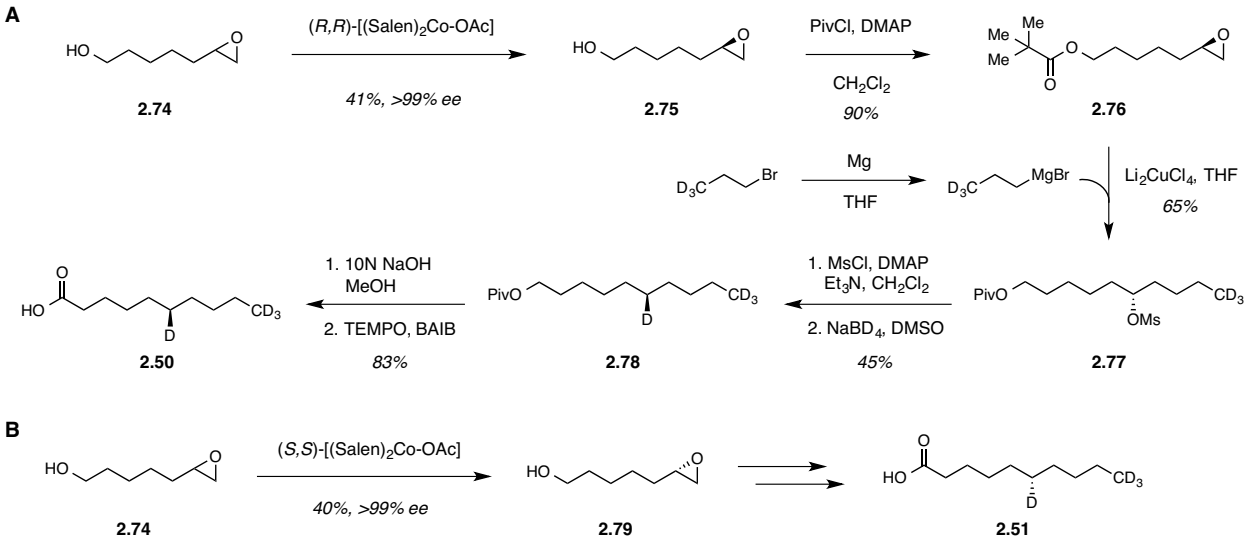
***D_3*-1,3-dimethoxy-5-(2-methylundec-1-en-1yl)benzene (2.73):** *d_3*-1-iodo-2-methylundec-1-ene **2.72** (123 mg, 0.414 mmol) was dissolved in THF (3 mL) and water (1 mL). To the reaction mixture, tetrakis(triphenylphosphine) palladium (48 mg, 0.041 mmol), silver oxide (24 mg, 0.103 mmol) and (3,5-dimethyoxyphenyl)boronic acid (87 mg, 0.476 mmol) were added. The reaction mixture was stirred at room temperature overnight. The reaction was diluted in 1:1 hexanes/diethyl ether (20 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 0-10% ethyl acetate in hexanes to afford *d_3*-1,3-dimethoxy-5-(2-methylundec-1-en-1yl)benzene **2.73** as a mixture of *E* and *Z* alkenes (60 mg, 0.195 mmol, 47% yield). TLC: R_f = 0.70 (5:1 hexanes/ethyl acetate). ^{1}H-NMR (500 MHz; $CDCl_3$): ***E* alkene:** δ 6.42 (s, 2H), 6.38 (s, 1H), 6.23 (s, 1H), 3.80 (s, 3H), 2.16 (t, *J* = 7.4 Hz, 2H), 1.51 (m, 2H), 1.34-1.27 (m, 12H), 0.90 (t, *J* = 7.0 Hz, 3H); ***Z* alkene:** δ 6.42 (s, 2H), 6.38 (s, 1H), 6.34 (s, 1H), 3.80 (s, 3H), 2.25 (t, *J* = 8.0 Hz, 2H), 1.51 (m, 2H), 1.34-1.27 (m, 12H), 0.90 (t, *J* = 7.0 Hz, 3H).



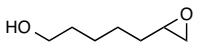
2.49

***D*$_3$-Resorcinol (2.49):** *D*$_3$-1,3-dimethoxy-5-(2-methylundec-1-en-1yl)benzene **2.73** (60 mg, 0.20 mmol) was dissolved in anhydrous ethyl acetate (5 mL), and palladium on carbon (41.5 mg, 0.039 mmol) was added to the reaction flask. The reaction mixture was stirred under an atmosphere of hydrogen (1 ATM) overnight. The reaction mixture was passed through a plug of celite and the filtrate was concentrated *in vacuo*. To the crude residue (57 mg, 0.184 mmol) dissolved in dichloromethane (2 mL) and cooled to 0 ºC, a solution of boron tribromide (1.0 M in dichloromethane, 1.84 mL, 1.84 mmol) was added to the flask dropwise over 5 min. The reaction mixture was warmed to room temperature and stirred overnight. The reaction mixture was then cooled to 0 ºC and quenched with saturated aqueous solution of potassium carbonate (10 mL) then extracted with dichloromethane (3 × 10 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo.* The crude product was purified by flash chromatography, eluting at 50-60% ethyl acetate in hexanes to afford the *d*$_3$-resorcinol **2.49** (46 mg, 0.16 mmol, 84% yield). TLC: R_f = 0.70 (5:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{18}H_{26}D_3O_2^-$ $[M–H]^-$, 280.3461; found 280.3463. ^{1}H-NMR (400 MHz, $CDCl_3$) δ: 6.22 (d, 2H, *J* = 2.2, aromatic C**H**), 6.18 (d, 1H, *J* = 2.2, aromatic C**H**), 5.28 (s, 2H, O**H**), 2.51 (dd, 1H, *J* = 13.4, 6.1, benzylic C**H**$_2$), 2.19 (dd, 1H, *J* = 13.6, 8.5, benzylic C**H**$_2$), 1.63 (m, 1H, C**H**), 1.23-1.35 (m, 15H, C**H**$_2$), 1.11 (m, 1H, C**H**$_2$), 0.88 (t, 3H, *J* = 6.8, CH_2C**H**$_3$). ^{13}C-NMR (125 MHz, CDCl3) δ: 156.3, 145.1, 108.9, 100.2, 43.5, 36.7, 34.5, 31.9, 29.9, 29.7, 29.6, 29.35, 27.1, 22.7, 19.7 (m), 14.1. (Note: ^{13}C peaks of the carbon atom bonded to the deuterium atoms is broad multiplet, which is difficult to identify).

2.4.18. Synthesis of stereoselectively labeled d_4-decanoic acids

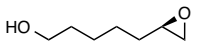


Scheme 2.26: Synthesis of stereospecifically-labeled d_4-decanoic acids (**2.50** and **2.51**) for feeding experiments.



2.74

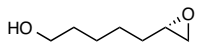
Synthesis of 5-(oxiran-2-yl)pentan-1-ol (**2.74**) was performed following a previously reported procedure.[37]



2.75

(*R*)-5-(Oxiran-2-yl)pentan-1-ol (2.75): (*R,R*)-[Co(salen)] complex (210 mg, 0.35 mmol) and acetic acid (40 µL, 0.69 mmol) were dissolved in toluene (3 mL). The reaction mixture was stirred at room temperature for 30 min and then concentrated *in vacuo*. To the dried (*R,R*)-[Co(salen)(OAc)] catalyst, racemic 5-(oxiran-2-yl)pentan-1-ol **2.74** (1.5 g, 11.5 mmol) was added. To the reaction mixture cooled to 0 ºC and water (125 µL, 6.9 mmol) was added. The reaction mixture was warmed to room temperature and stirred for an additional 14 h. Silica gel was added to the flask and the solvent was removed *in vacuo.* The resulting residue was dry loaded onto a column for purification by flash chromatography, eluting with 40-50% ethyl acetate in hexanes to afford (*R*)-5-(oxiran-2-yl)pentan-1-ol **2.75** as a colorless oil (615

mg, 4.7 mmol, 41%). TLC: R_f = 0.29 (1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_7H_{15}O_2^+$ $[M+H]^+$, 131.1067; found 131.1066. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.63 (t, 2H, *J* = 6.6 Hz, CH_2OH), 2.90 (m, 1H, OCH), 2.74 (dd, 1H, *J* = 5.0, 4.0 Hz, OCH_2), 2.46 (dd, 1H, *J* = 5.0, 2.8 Hz, OCH_2), 1.60-1.38 (m, 8H, CH_2). ^{13}C NMR (125 MHz; $CDCl_3$): δ 62.9, 52.4, 47.2, 32.7, 32.5, 25.9, 25.7.



2.79

(*S*)-5-(Oxiran-2-yl)pentan-1-ol (2.79): Synthesis was performed following the procedure described for the preparation of (*R*)-5-(oxiran-2-yl)pentan-1-ol **2.75**, except using (*S,S*)-[Co(salen)] complex as the catalyst to obtain (*S*)-5-(oxiran-2-yl)pentan-1-ol **2.79** as a colorless oil (594 mg, 4.6 mmol, 40%). TLC: R_f = 0.29 (1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_7H_{15}O_2^+$ $[M+H]^+$, 131.1067; found 131.1067. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.63 (t, 2H, *J* = 6.6 Hz, CH_2OH), 2.90 (m, 1H, OCH), 2.74 (dd, 1H, *J* = 5.0, 4.0 Hz, OCH_2), 2.46 (dd, 1H, *J* = 5.0, 2.8 Hz, OCH_2), 1.60-1.38 (m, 8H, CH_2). ^{13}C NMR (125 MHz; $CDCl_3$): δ 62.9, 52.4, 47.2, 32.7, 32.5, 25.9, 25.7.

Enantiomeric excess analysis of (*R*)- and (*S*)-(oxiran-2-yl)pentan-1-ol (2.75 and 2.79): (*R*)- **2.75** or (*S*)-(oxiran-2-yl)pentan-1-ol **2.79** (5.0 mg, 0.040 mmol) and 2-naphthalenethiol (12 mg, 0.080 mmol) were dissolved in anhydrous methanol (1 mL) at room temperature. Triethylamine (10 μL, 0.070 mmol) was added and the reaction mixture was stirred at room temperature for 1 h. The derivatized epoxide was purified by flash chromatography, eluting with 10% ethyl acetate in hexanes, and then re-dissolved in 1:9 *i*PrOH/hexanes (1 mg/mL). The *ee* of the product was determined to be >99% by chiral HPLC analysis (ChiralPak AD-H, 5 μL injection volume, 10% isopropanol in hexanes, 1 mL/min, 254 nm).

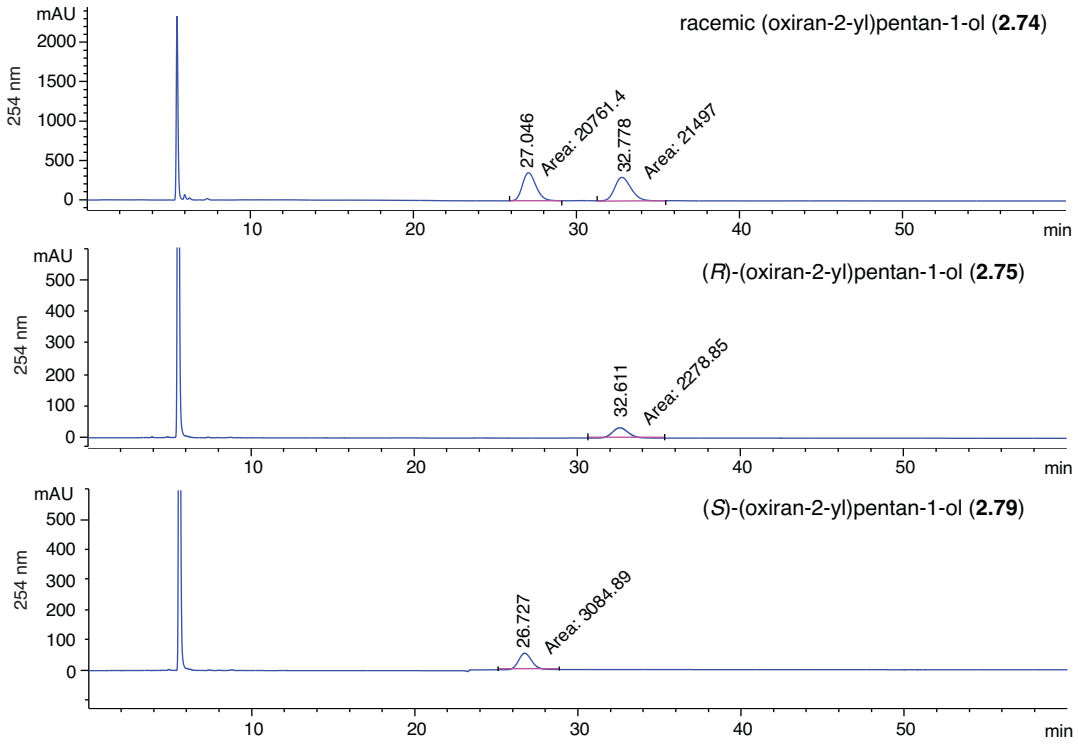
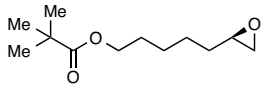


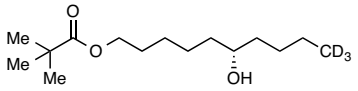
Figure 2.26: Chiral HPLC *ee* analysis of the chiral epoxides.



2.76

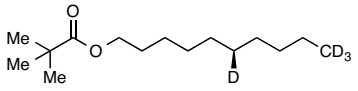
(*R*)-5-(oxiran-2-yl)pentyl pivalate (2.76): To a solution of 4-dimethylaminopyridine (1.4 g, 11.5 mmol) in dichloromethane (3 mL) cooled to 0 ºC, pivaloyl chloride (740 µL, 6.0 mmol) was added. To the reaction mixture, (*R*)-(oxiran-2-yl)pentan-1-ol (600 mg, 4.6 mmol) in dichloromethane (3 mL) was added dropwise over 5 min. The reaction mixture was warmed to room temperature and stirred for 1.5 h. The reaction mixture was quenched with 1 M aqueous hydrochloric acid (10 mL) and extracted with dichloromethane (3 □ 10 mL). The combined organic layers were washed with 1 M HCl (20 mL), and brine (30 mL), dried over anhydrous sodium sulfate, filtered and concentrated *in vacuo*. The crude residue was purified by flash chromatography, eluting at 10% ethyl acetate in hexanes to afford (*R*)-5-(oxiran-2-yl)pentyl pivalate **2.76** as a clear oil (890 mg, 4.1 mmol, 90%). TLC: R_f = 0.83 (1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{12}H_{23}O_3^+$ $[M+H]^+$, 215.1642; found 215.1653. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.98 (t, 2H, *J* = 6.6 Hz, COOC**H**$_2$), 2.82 (m, 1H, OC**H**), 2.66 (dd, 1H, *J* = 5.0, 4.0 Hz, OC**H**$_2$), 2.38 (dd,

1H, *J* = 5.1, 2.7 Hz, OC**H**$_2$), 1.57 (m, 2H, C**H**$_2$), 1.49-1.34 (m, 6H, C**H**$_2$), 1.11 (s, 9H, CC**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 178.3, 64.0, 51.9, 46.7, 38.5, 32.1, 28.3, 27.0, 25.6, 25.4.



2.77

***D*$_3$-(*S*)-6-Hydroxydecyl pivalate (2.77):** (*R*)-5-(Oxiran-2-yl)pentyl pivalate **2.76** (500 mg, 2.3 mmol) was dissolved in THF (6 mL) and the resulting solution was cooled to –30 ºC. Then, Li_2CuCl_4 solution (0.1 M in THF, 2.33 mL) was added dropwise to the reaction mixture over 5 min, and the reaction mixture was stirred at –30 ºC for an additional 15 min. To the reaction mixture, *n*-propylmagnesium bromide solution (0.5 M, 5.6 mL, 2.8 mmol) was added dropwise over 15 min, and the reaction mixture was stirred at –30 ºC for 1 h. The reaction mixture was diluted with diethyl ether (20 mL) and quenched with saturated aqueous solution of ammonium chloride (20 mL). The quenched reaction mixture was extracted with diethyl ether (3 x 20 mL). The combined organic layers were washed with water (40 mL) and brine (40 mL), dried over anhydrous sodium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 10-20% ethyl acetate in hexanes to afford *d*$_3$-(*S*)-6-hydroxydecyl pivalate **2.77** as a colorless oil (390 mg, 1.5 mmol, 65%). TLC: R_f = 0.23 (4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{15}H_{28}D_3O_3^+$ $[M+H]^+$, 262.2456; found 262.2469. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 4.00 (t, 2H, *J* = 6.6 Hz, COOC**H**$_2$), 3.53 (m, 1H, C**H**OH), 1.79 (s, 1H, O**H**), 1.59 (m, 2H, C**H**$_2$), 1.44-1.22 (m, 12H, C**H**$_2$), 1.14 (s, 9H, CC**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 178.7, 71.8, 64.4, 38.8, 37.4, 37.3, 28.7, 27.8, 27.2, 26.1, 25.3, 22.6, 13.2 (m). (Note: ^{13}C peaks of the carbon atom bonded to the deuterium atoms is broad multiplet).

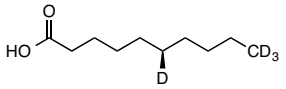


2.78

***D*$_4$-(*R*)-6-Decyl pivalate (2.78):** *d*$_3$-(*S*)-6-hydroxydecyl pivalate **2.77** (400 mg, 1.5 mmol) and 4-dimethylaminopyridine (47 mg, 0.38 mmol) were dissolved in dichloromethane (5 mL). To the reaction

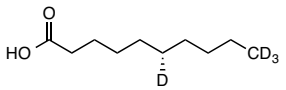
mixture was cooled to 0 ºC, triethylamine (1.1 mL, 7.7 mmol) and mesyl chloride (130 µL, 1.6 mmol) were added. The reaction mixture was stirred for 30 min at 0 ºC and then quenched with 1 M aqueous potassium hydroxide (5 mL). The quenched reaction mixture was acidified with 1 M aqueous hydrochloric acid (10 mL) and extracted with dichloromethane (3 x 10 mL). The combined organic layers were washed with water (30 mL), dried over sodium sulfate, filtered and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 10-15% ethyl acetate in hexanes to afford the mesylate (330 mg, 0.98 mmol). TLC: R_f = 0.38 (4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{16}H_{30}D_3O_5S^+$ $[M+H]^+$, 340.2232; found 340.2205. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 4.62 (m, 1H, *J* = 6.1 Hz, SOC**H**), 3.97 (t, 2H, *J* = 6.6 Hz, COOC**H**$_2$), 2.93 (s, 3H, SC**H**$_3$), 1.64-1.54 (m, 6H, C**H**$_2$), 1.38-1.23 (m, 8H, C**H**$_2$), 1.11 (s, 9H, CC**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 178.4, 83.8, 64.1, 38.7, 38.6, 34.3, 34.1, 28.4, 27.4, 27.0, 25.7, 24.5, 22.1, 13.1 (m). (Note: ^{13}C peaks of the carbon atom bonded to the deuterium atoms is broad multiplet).

To the mesylate (150 mg, 0.44 mmol) dissolved in anhydrous DMSO (10 mL), sodium borodeuteride (190 mg, 4.4 mmol) was added. The reaction mixture was warmed to 80 ºC for 2 h, then cooled to room temperature and then quenched with acetic acid (0.2 mL). The reaction mixture was diluted with saturated aqueous solution of sodium bicarbonate (10 mL) and extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with sodium bicarbonate (20 mL) and brine (20 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to afford d_4-(*R*)-6-decyl pivalate **2.78** as a colorless oil (160 mg, 0.31 mmol, 71%). TLC: R_f = 0.76 (4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{15}H_{27}D_4O_2^+$ $[M+H]^+$, 247.2570; found 247.2578. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 4.04 (t, 2H, *J* = 6.6 Hz, COOC**H**$_2$), 1.63-1.58 (m, 2H, C**H**$_2$), 1.35-1.22 (m, 12H, C**H**$_2$), 1.17 (s, 9H, CC**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 178.7, 64.6, 38.8, 31.9, 29.6 (2), 29.33 (t), 29.23, 28.8, 27.3, 26.0, 22.5, 13.3 (m). (Note: ^{13}C peaks of the carbon atom bonded to the deuterium atoms is broad multiplet).



2.50

***D*$_4$-(*R*)-Decanoic acid (2.50):** *D*$_4$-(*R*)-6-decyl pivalate **2.78** (160 mg, 0.65 mmol) was dissolved in methanol (10 mL) and 10 M aqueous sodium hydroxide (3 mL) was added. The reaction mixture was stirred overnight at room temperature. The reaction mixture was quenched with saturated aqueous solution of ammonium chloride (20 mL) and extracted with diethyl ether (3 x 30 mL). The combined organic layers were washed with saturated aqueous solution of sodium bicarbonate (50 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The resulting crude product was dissolved in 1:1 water/acetonitrile (2 mL). TEMPO (20 mg, 0.13 mmol) and (diacetoxyiodo)benzene (500 mg, 1.6 mmol) were added, and the reaction mixture was stirred overnight. The reaction mixture was quenched with saturated aqueous solution of sodium thiosulfate (10 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layers were dried over sodium sulfate, filtered, and concentrated *in vacuo*. The crude residue was purified by flash chromatography, eluting with 20-30% ethyl acetate in hexanes to afford chiral *d*$_4$-(*R*)-decanoic acid **2.50** as a colorless oil (95 mg, 0.54 mmol, 83%). TLC: R_f = 0.5 (1:1 hexanes/ethyl acetate). ^{1}H-NMR (500 MHz; $CDCl_3$): δ 2.34 (t, 2H, *J* = 7.5 Hz, COC**H**$_2$), 1.63 (m, 2H, C**H**$_2$), 1.36-1.23 (m, 10H, C**H**$_2$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 180.6, 34.3, 31.9, 29.29, 29.27, 29.16, 29.10 (m), 24.8, 22.5. 13.1 (m). (Note: ^{13}C peaks of the carbon atom bonded to the deuterium atoms is broad multiplet).



2.51

***D*$_4$-(*S*)-Decanoic acid (2.51):** Synthesis was performed following the same procedure as the synthesis of *d*$_4$-(*R*)-decanoic acid **2.50**, except using *d*$_4$-(*S*)-6-decyl pivalate (160 mg, 0.65 mmol) as the starting material to obtain *d*$_4$-(*S*)-decanoic acid **2.51** (100 mg, 0.56 mmol, 87%). ^{1}H and ^{13}C NMR spectra of *d*$_4$-(*S*)-decanoic acid **2.51** matched those of *d*$_4$-(*R*)-decanoic acid **2.50**.

2.5. References

(1) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P. Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J. Am. Chem. Soc.* **2012**, *134*, 18518.

(2) Bobzin, S. C.; Moore, R. E. Biosynthetic origin of [7.7]paracyclophanes from cyanobacteria. *Tetrahedron* **1993**, *49*, 7615.

(3) Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* **2008**, *25*, 845.

(4) Haapalainen, A. M.; Merilainen, G.; Wierenga, R. K. The thiolase superfamily: Condensing enzymes with diverse reaction specificities. *Trends Biochem. Sci* **2006**, *31*, 64.

(5) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **2008**, *26*, 225.

(6) Chang, Z. X.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J. Y.; Sherman, D. H.; Gerwick, W. H. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2004**, *67*, 1356.

(7) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium lyngbya majuscula. *Chem. Biol.* **2004**, *11*, 817.

(8) Jones, A. C.; Monroe, E. A.; Eisman, E. B.; Gerwick, L.; Sherman, D. H.; Gerwick, W. H. The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat. Prod. Rep.* **2010**, *27*, 1048.

(9) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E. Structures of cylindrocyclophanes A-F. *Tetrahedron* **1992**, *48*, 3001.

(10) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. Blast+: Architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421.

(11) Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S. N.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z. X.; Yamashita, R. A.; Zhang, D. C.; Zheng, C. J.; Bryant, S. H. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, 222.

(12) Bachmann, B. O.; Ravel, J. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Method Enzymol* **2009**, *458*, 181.

(13) Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489.

(14) Austin, M. B.; Noel, A. J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **2003**, *20*, 79.

(15) La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. Manipulation of carrier proteins in antibiotic biosynthesis. *Chem. Biol.* **2004**, *11*, 195.

(16) Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **2006**, *1*, 280.

(17) Chen, J. L.; Moore, R. E.; Patterson, G. M. L. Structures of nostocyclophanes A-D. *J. Org. Chem.* **1991**, *56*, 4360.

(18) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. Carbamidocyclophanes A-E, chlorinated paracyclophanes with cytotoxic and antibiotic activity from the Vietnamese cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2007**, *70*, 499.

(19) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. Cylindrocyclophanes with proteasome inhibitory activity from the cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2010**, *73*, 1529.

(20) Luo, S. W.; Kang, H. S.; Krunic, A.; Chlipala, G. E.; Cai, G. P.; Chen, W. L.; Franzblau, S. G.; Swanson, S. M.; Orjala, J. Carbamidocyclophanes F and G with anti-*Mycobacterium tuberculosis* activity from the cultured freshwater cyanobacterium *Nostoc* sp. *Tetrahedron Lett.* **2014**, *55*, 686.

(21) Preisitsch, M.; Harmrolfs, K.; Pham, H. T. L.; Heiden, S. E.; Fuessel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H. J.; Müller, R.; Mundt, S. Anti-MRSA-acting carbamidocyclophanes H-L from the Vietnamese cyanobacterium *Nostoc* sp. CAVN2. *J. Antibiot.* **2015**, *68*, 600.

(22) Trivedi, O. A.; Arora, P.; Sridharan, V.; Tickoo, R.; Mohanty, D.; Gokhale, R. S. Enzymic activation and transfer of fatty acids and acyl-adenylates in mycobacteria. *Nature* **2004**, *430*, 810.

(23) Arora, P.; Goyal, A.; Natarajan, V. T.; Rajakumara, E.; Verma, P.; Gupta, R.; Yousuf, M.; Trivedi, O. A.; Mohanty, D.; Tyagi, A.; Sankaranarayanan, R.; Gokhale, R. S. Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nat. Chem. Biol.* **2009**, *5*, 166.

(24) Hansen, D. B.; Bumpus, S. B.; Aron, Z. D.; Kelleher, N. L.; Walsh, C. T. The loading module of mycosubtilin: An adenylation domain with fatty acid selectivity. *J. Am. Chem. Soc.* **2007**, *129*, 6366.

(25) Austin, M. B.; Saito, T.; Bowman, M. E.; Haydock, S.; Kato, A.; Moore, B. S.; Kay, R. R.; Noel, J. P. Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid - type III polyketide synthase. *Nat. Chem. Biol.* **2006**, *2*, 494.

(26) Miyanaga, A.; Funa, N.; Awakawa, T.; Horinouchi, S. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 871.

(27) Song, L. J.; Barona-Gomez, F.; Corre, C.; Xiang, L. K.; Udwary, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L. Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by *Streptomyces coelicolor* genome mining. *J. Am. Chem. Soc.* **2006**, *128*, 14754.

(28) Chemler, J. A.; Buchholz, T. J.; Geders, T. W.; Akey, D. L.; Rath, C. M.; Chlipala, G. E.; Smith, J. L.; Sherman, D. H. Biochemical and structural characterization of germicidin synthase: Analysis of a

type III polyketide synthase that employs acyl-ACP as a starter unit donor. *J. Am. Chem. Soc.* **2012**, *134*, 7359.

(29) Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in *Bacillus subtilis*. *J. Bacteriol.* **2009**, *191*, 4916.

(30) von Berlepsch, S.; Kunz, H. H.; Brodesser, S.; Fink, P.; Marin, K.; Flugge, U. I.; Gierth, M. The acyl-acyl carrier protein synthetase from *Synechocystis* sp. PCC 6803 mediates fatty acid import. *Plant Physiol.* **2012**, *159*, 606.

(31) Smith, A. B.; Kozmin, S. A.; Paone, D. Total synthesis of (–)-cylindrocyclophane F. *J. Am. Chem. Soc.* **1999**, *121*, 7423.

(32) Withall, D. M.; Haynes, S. W.; Challis, G. L. Stereochemistry and mechanism of undecylprodigiosin oxidative carbocyclization to streptorubin B by the Rieske oxygenase RedG. *J. Am. Chem. Soc.* **2015**, *137*, 7889.

(33) Gelb, M. H.; Heimbrook, D. C.; Malkonen, P.; Sligar, S. G. Stereochemistry and deuterium-isotope effects in camphor hydroxylation by the cytochrome P450cam monoxygenase system. *Biochemistry* **1982**, *21*, 370.

(34) Nishikado, H.; Nakatsuji, H.; Ueno, K.; Nagase, R.; Tanabe, Y. Mild, efficient, and robust method for stereocomplementary iron-catalyzed cross-coupling using (*E*)- and (*Z*)-enol tosylates. *Synlett* **2010**, 2087.

(35) Mino, T.; Masuda, S.; Nishio, M.; Yamashita, M. Synthesis of lactones by Baeyer–Villiger oxidation with magnesium monoperphthalate hexahydrate. *J. Org. Chem.* **1997**, *62*, 2633.

(36) Holmquist, H. E.; Rothrock, H. S.; Theobald, C. W.; Englund, B. E. Some decomposition and rearrangement products of decahydronaphthalene hydroperoxide. *J. Am. Chem. Soc.* **1956**, *78*, 5339.

(37) Mihailovic, M. L.; Marinkovic, D. The formation of cyclic ethers from olefinic alcohols. The oxidative cyclization of some open-chain unsaturated alcohols by means of organic peracids. *Croat. Chem. Acta* **1986**, *59*, 109.

Chapter 3. Investigation of the assembly line termination in the cylindrocyclophane biosynthesis

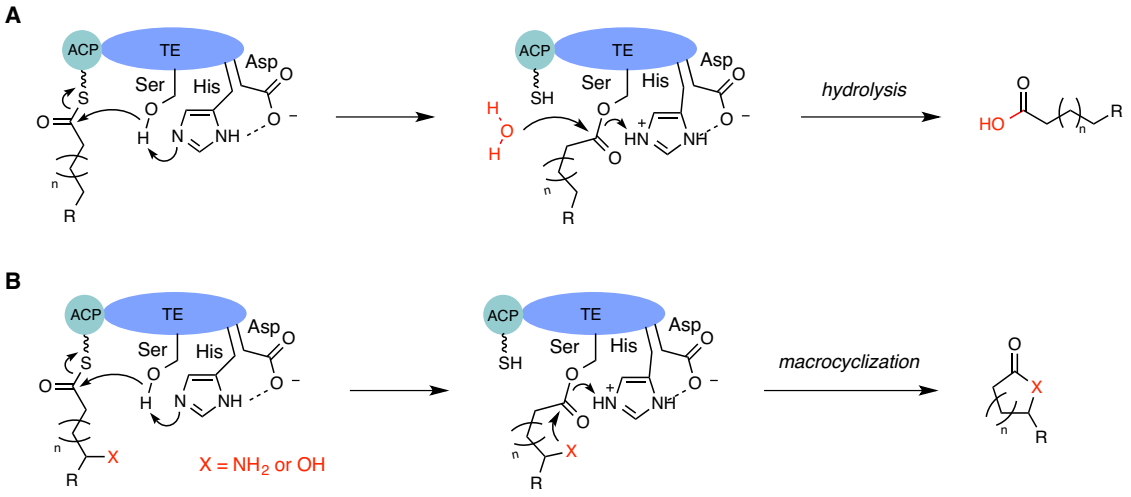
Parts of this chapter are adopted from previous publications.[1]

3.1. Introduction

Polyketides are a large group of secondary metabolites that display a wide range of important biological activities.[2-4] Many polyketides are synthesized by type I polyketide synthases (PKSs), which generate structural diversity and complexity by employing combinations of domains and modules organized into enzymatic assembly lines.[5] One important source of structural diversity in polyketide biosynthesis is the assembly line termination step.[6] At the end of an assembly line, a nascent polyketide product must be released from the PKS by a terminating enzyme. Three major types of termination logic are used by type I PKS assembly lines: release by a C-terminal thioesterase (TE) domain, release by an alternate C-terminal domain, and release by a freestanding enzyme *in trans*.[6]

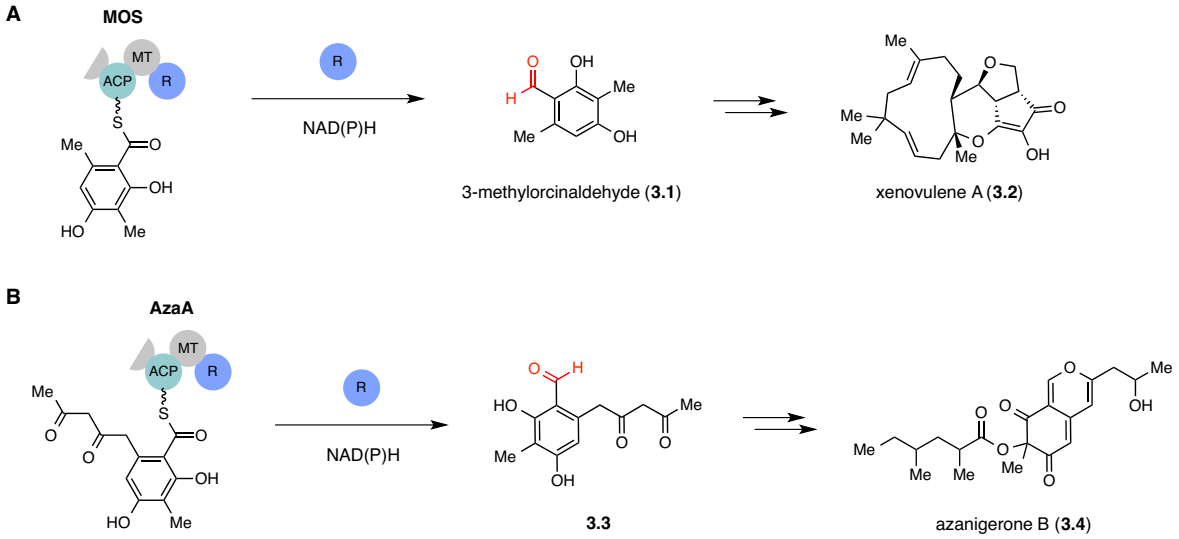
3.1.1. Termination of PKS assembly lines

The canonical and most well studied termination logic is the C-terminal TE domain-mediated assembly line termination (**Scheme 3.1**). TE domains that catalyze assembly line terminations are referred to as type I TEs and are typically located at the C-terminus of the final module of type I PKSs.[6] TE domains release the nascent polyketide product from assembly lines by catalyzing either hydrolysis (**Scheme 3.1A**) or macrocyclization (**Scheme 3.1B**) using a conserved Ser-His-Asp catalytic triad.

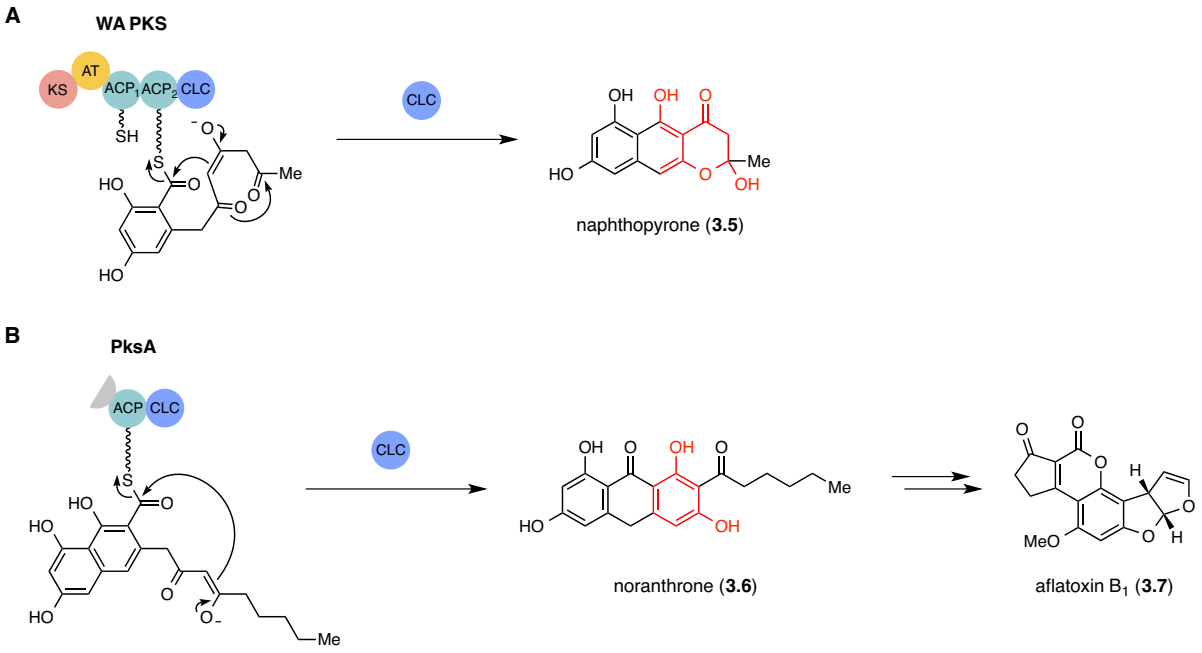


Scheme 3.1: Typical *in cis* polyketide product release catalyzed by the C-terminal TE domains using the Ser-His-Asp triad. **A**) Hydrolysis of the polyketide to form free acids. **B**) Macrocyclization with the intramolecular nucleophile to form macrolactones or macrolactams.

Another common termination logic found in PKS assembly lines involves product release by C-terminal domains other than TEs.[6] For example, reductase (R) domains catalyze reductive release using NAD(P)H cofactors in multiple PKS assembly lines, including those in xenovulene (**Scheme 3.2A**)[7] and azanigerone (**Scheme 3.2B**)[8] biosyntheses. Product release by R domains results in the formation of aldehydes through a two-electron reduction or primary alcohols through a four-electron reduction. Furthermore, a special type of TE domain called a Claisen-like cyclase (CLC) catalyzes Claisen-type condensations to form a new ring system in the biosyntheses of the fungal polyketides, such as naphthopyrone (**Scheme 3.3A**)[9] and aflatoxin (**Scheme 3.3B**).[10] Product release catalyzed by domains that are integrated into PKS assembly lines are considered to use *in cis* termination logic.



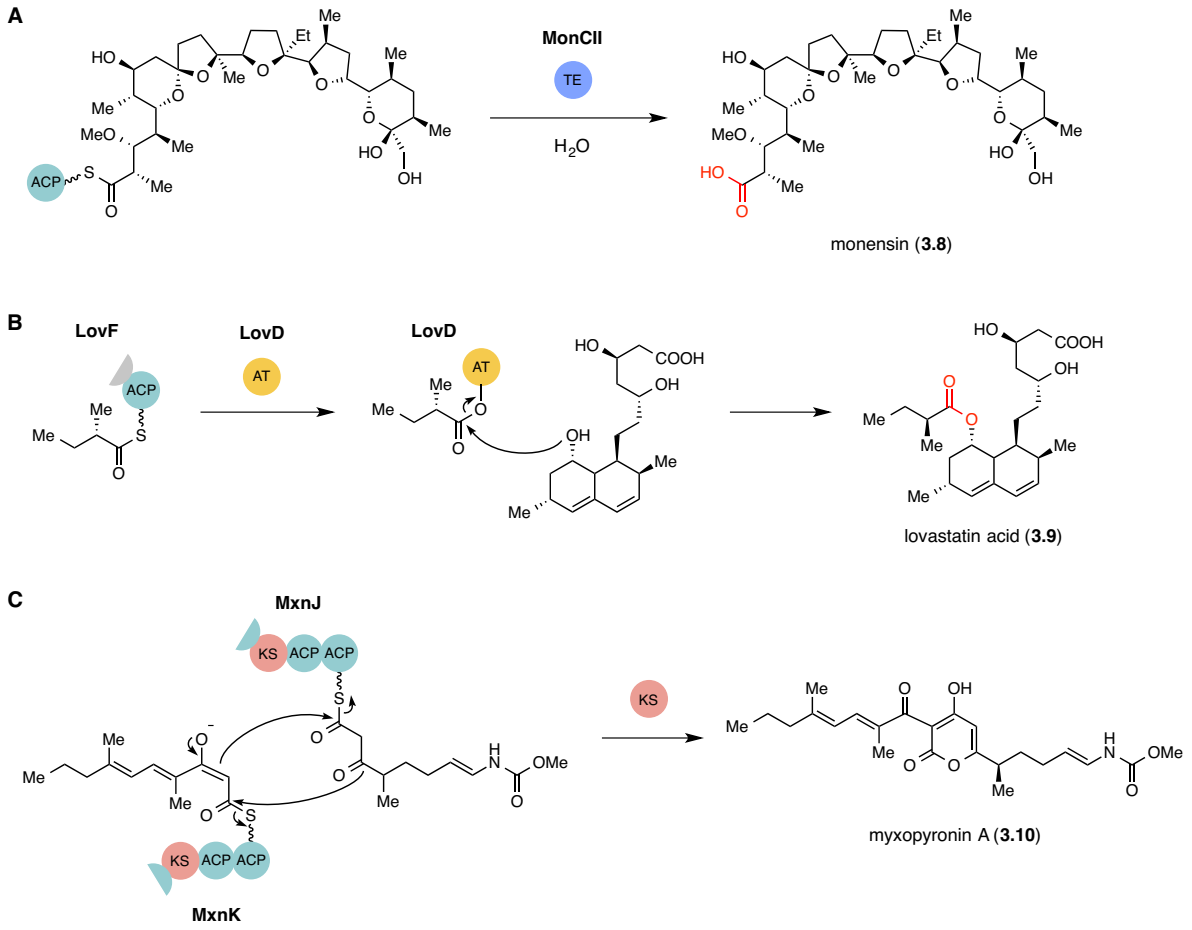
Scheme 3.2: *In cis* reductive release by R domains of polyketide products by PKSs. **A**) Reductive release in xenovulene biosynthesis. **B**) Reductive release in azanigerone biosynthesis.



Scheme 3.3: *In cis* PKS product release via Claisen-cyclization. **A**) Product release catalyzed by the CLC domain in naphthopyrone biosynthesis. **B**) Product release catalyzed by the CLC domain in aflatoxin biosynthesis.

The third type of termination logic seen in PKS assembly lines involves *in trans* product release catalyzed by discrete enzymes that are not integrated into the enzymatic assembly line.[6] PKSs that use *in trans* product release typically lack a C-terminal domain for product release. Examples of *in trans* assembly

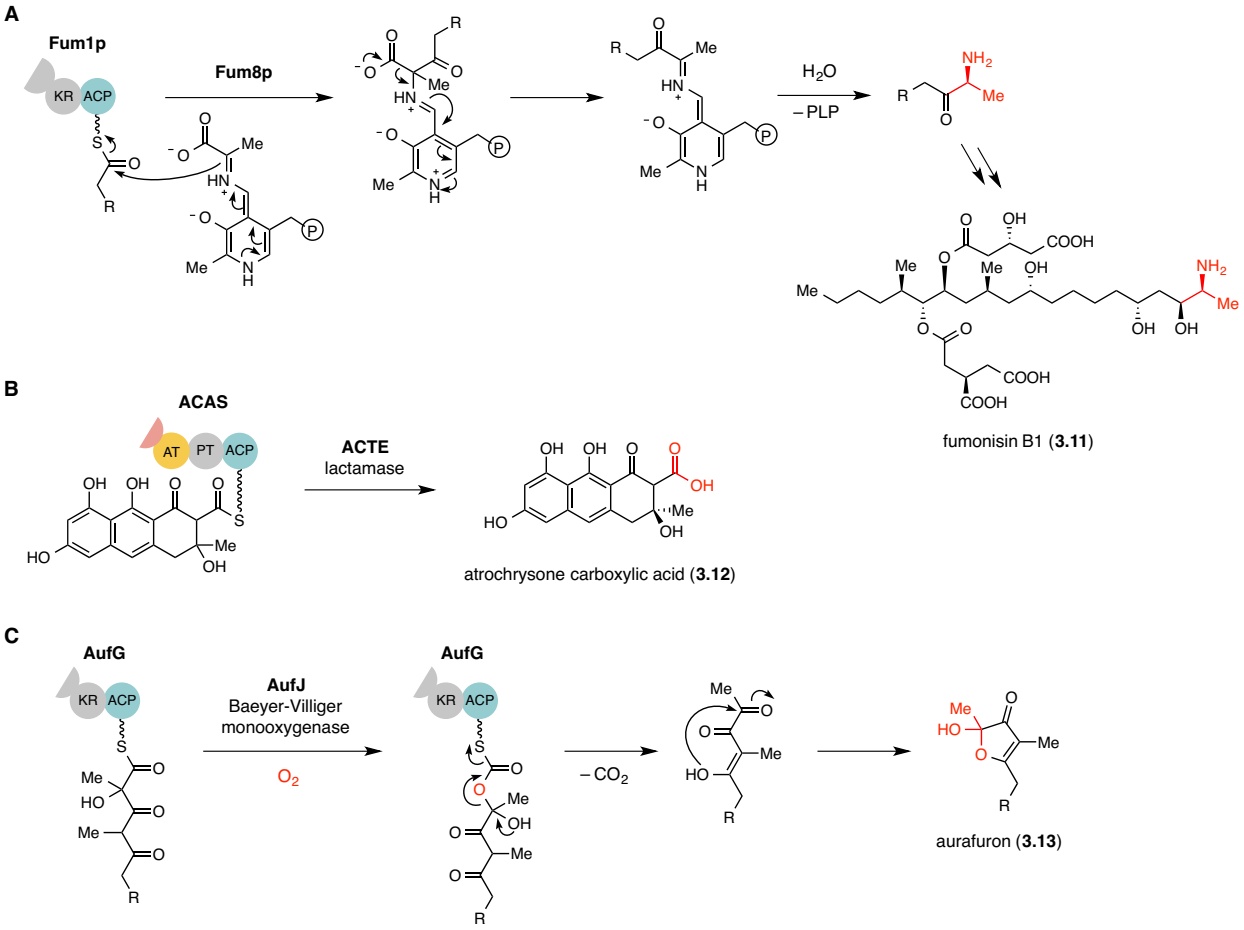
line termination include hydrolysis catalyzed by a freestanding TE in the biosynthesis of polyether ionophores (**Scheme 3.4A**),[11-13] and acyl transfer to an alcohol reaction partner catalyzed by a freestanding AT in the biosynthesis of lovastatin (**Scheme 3.4B**).[14] Additionally, a freestanding ketosynthase has recently been shown to condense two nascent polyketides, each tethered to two separate type I modular PKSs, to form a pyrone ring in myxopyronin biosynthesis (**Scheme 3.4C**).[15]



Scheme 3.4: *In trans* product release in type I modular PKS assembly lines. **A**) Hydrolysis catalyzed by a freestanding TE in monensin biosynthesis. **B**) Transfer reaction catalyzed by a freestanding AT in lovastatin biosynthesis. **C**) Condensation catalyzed by a freestanding KS in myxopyronin biosynthesis.

In trans termination logic is often found in type I iterative PKS systems.[6] Enzymes that catalyze unique product release steps in type I iterative PKS assembly lines include a pyridoxal 5'-phosphate (PLP)-dependent enzyme in fumonisin biosynthesis (**Scheme 3.5A**),[16] a lactamase in atrochrysone carboxylic acid biosynthesis (**Scheme 3.5B**),[17] and a Baeyer-Villiger monooxygenase in aurafuron biosynthesis

(**Scheme 3.5C**).[18] The variety of enzymes that participate in product release further illustrates the important contribution of the termination step to the structural diversity of polyketide biosynthetic products.



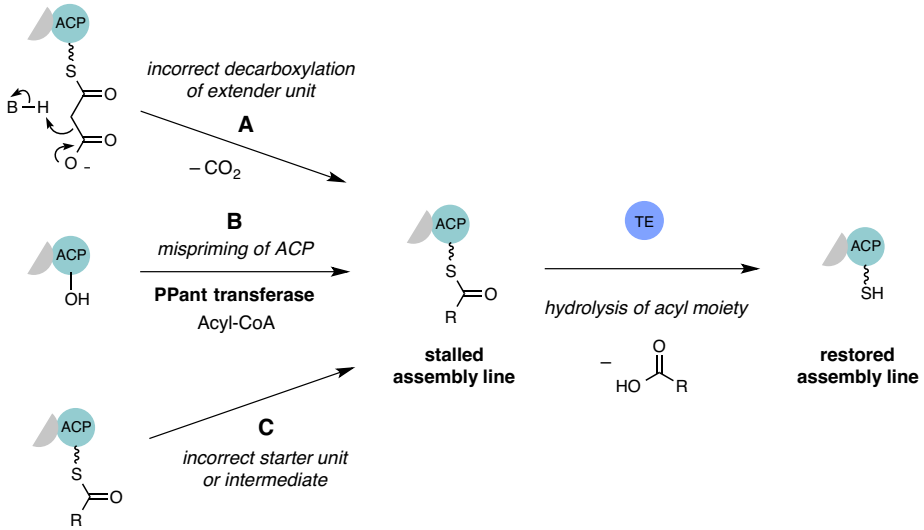
Scheme 3.5: Unusual product release in type I iterative PKS assembly lines. **A**) Product release catalyzed by a PLP-dependent enzyme in fumonisin biosynthesis. **B**) Hydrolysis catalyzed by a lactamase in atrochrysone carboxylic acid biosynthesis. **C**) Product release catalyzed by a Baeyer-Villiger monooxygenase in aurafuron biosynthesis.

3.1.2. Types II thioesterases play editing functions

Unlike type I TEs, which are located at the C-terminus of PKS or NRPS enzymes to catalyze product release, type II TE are discrete enzymes that are also associated with biosynthetic pathways with assembly line enzymes.[19] While some type II TEs are involved in assembly line terminations, such as the TE in monensin (**3.8**, see **Scheme 3.4**) biosynthesis,[12] most do not play a role in product release.[20] Instead,

type II TEs are known to play editing functions that are beneficial to maintaining efficient product formation by assembly line enzymes.[20-25]

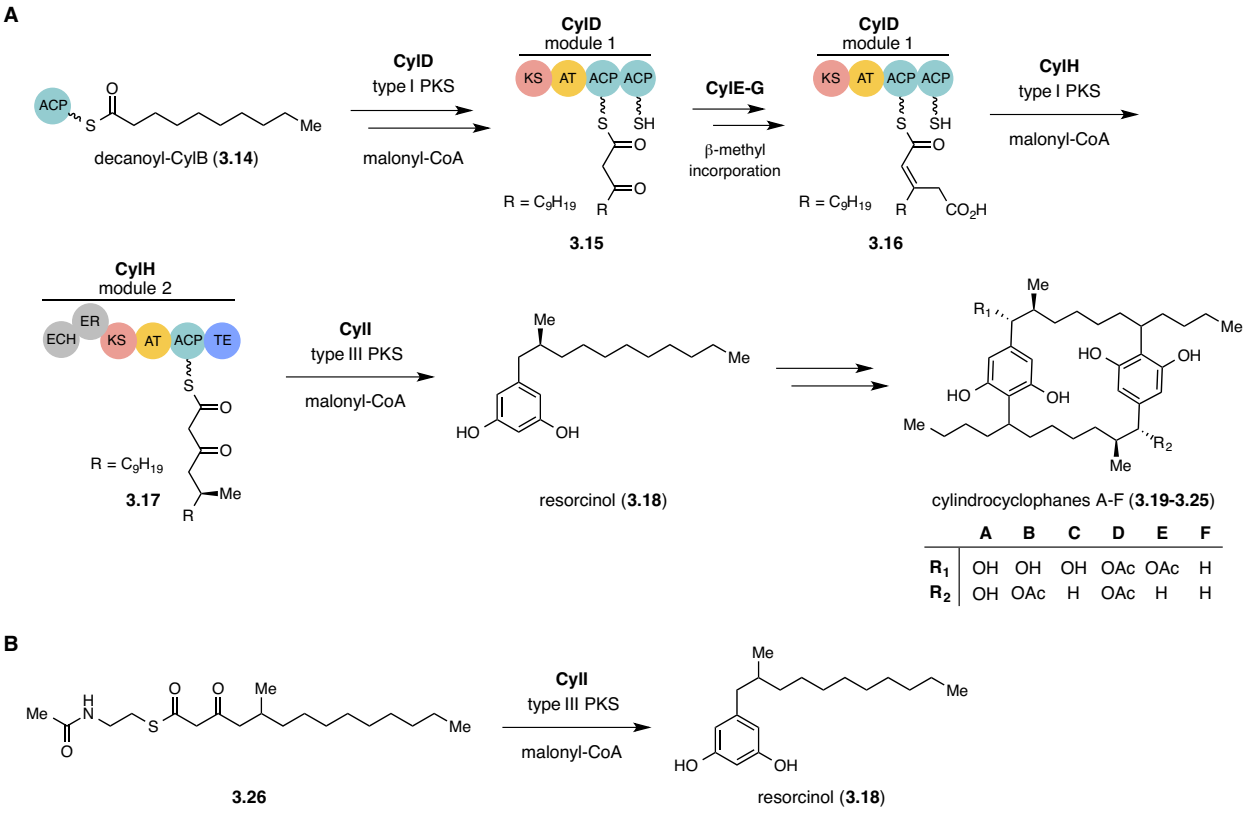
Editing functions catalyzed by type II TEs include hydrolytic cleavage of acyl residues that result from aberrant decarboxylation of extender units (**Scheme 3.6A**), mispriming of ACP domains by phosphopantetheinyl (ppant) transferase (**Scheme 3.6B**), or loading of incorrect starter units or intermediates (**Scheme 3.6C**) in the context of PKS assembly lines.[20,23] The editing function of type II TEs requires these enzymes to have a broad substrate scope, and consequently the rate of hydrolysis by type II TEs is relatively slow.[20] Because the production of PKS and NRPS assembly line enzymes is energetically demanding, the organisms that produce such pathways benefit from restoring the activity of the assembly line enzymes that have become blocked by nonreactive acyl moieties.[20] Although many PKS and NRPS assembly lines require type II TEs to maintain efficient product formation, in most cases, type II TEs are not essential for the production of polyketides and non-ribosomal peptides.[20]



Scheme 3.6: Type II TEs restore functional assembly lines by catalyzing removal of acyl moieties from stalled PKSs. **A**) Aberrant decarboxylation of extender unit, **B**) mispriming of *apo*ACP by PPant transferases, and **C**) loading of incorrect starter unit or intermediate to ACP domains result in accumulation of stalled assembly lines.

3.1.3. Assembly line termination in the cylindrocyclophane biosynthesis

The cylindrocyclophanes (**3.19-3.25**, see **Scheme 3.7A**) are a family of polyketides with a rare [7.7]paracyclophane scaffold produced by a cyanobacterial biosynthetic pathway containing a type I modular PKS assembly line.[26] Our initial investigation of the cylindrocyclophanes biosynthesis revealed that this pathway involved a unique assembly line termination mechanism.[26] The type I modular PKS assembly line (CylD–CylH) in the cylindrocyclophane biosynthesis has a C-terminal TE domain. Nevertheless, our initial *in vitro* characterization of the type III PKS CylI suggested that this freestanding enzyme was responsible for catalyzing the release of the nascent polyketide **3.17** (**Scheme 3.7A**). We performed an *in vitro* assay using a synthetic β-ketoacyl-SNAC substrate (**3.26**, see **Scheme 3.7B**) that mimicked the structure of the polyketide product **3.17** predicted to be covalently tethered to the ACP domain of the type I PKS CylH. Incubation of the β-ketoacyl-SNAC **3.26** and malonyl-CoA with CylI resulted in the formation of resorcinol **3.18**. This *in vitro* assay result indicated that type III PKS CylI was involved in assembly line product release in the cylindrocyclophane biosynthesis. While several type III PKSs are known to utilize acyl-ACP substrates tethered to fatty acid synthases,[27-29] to our knowledge, CylI is the first type III PKS predicted to use an acyl-ACP substrate generated by a type I modular PKS.[26]



Scheme 3.7: Type I PKS assembly line termination in the cylindrocyclophane biosynthesis. **A**) The termination of the type I PKS assembly line appears to be catalyzed by the type III PKS CylI. **B**) *In vitro* CylI assay with the synthetic -ketoacyl-SNAC substrate leads to resorcinol formation.

The direct use of -ketoacyl-SNAC substrate **3.26** by CylI suggested that the C-terminal TE domain of type I PKS CylH did not participate in product release. Although there are many examples of PKS assembly line termination catalyzed by discrete enzymes *in trans*, all previously characterized pathways with *in trans* termination logic lack C-terminal TE domains. The type I modular cylindrocyclophane PKS assembly line was therefore unique in possessing both a freestanding terminating enzyme (CylI) and a C-terminal TE domain on its final module (CylH). This is surprising, as the termination by a type III PKS should not require a type I TE. To better understand the unusual termination logic involved in cylindrocyclophane biosynthesis, we further investigated the activity of CylI, the final module of CylH, CylI, and the excised CylH TE domain.

3.2. Results and Discussions

3.2.1. Assessment of the role of the CylH TE domain in assembly line termination

To determine whether the CylH TE domain is catalytically active and has a role in assembly line termination, we cloned and heterologously expressed the final PKS module of CylH ($CylH_{PKS}$ = KS-AT-ACP-TE) and the excised CylH TE domain ($CylH_{TE}$). The $CylH_{TE}$ construct also included a 20 amino acid linker region at the N-terminus. Additionally, we generated inactive TE domain mutants by using site-directed mutagenesis to convert the catalytic serine residue to an alanine ($CylH_{PKS}$-S1201A and $CylH_{TE}$-S95A). CylH domain boundaries were determined using the PKS/NRPS analysis web server (http://nrps.igs.umaryland.edu/nrps/), and the catalytic serine residue was identified through multiple sequence alignment of the CylH TE domain with other previously characterized TE domains (**Figure 3.1**). C-His_6-tagged $CylH_{TE}$, $CylH_{TE}$-S95A, $CylH_{PKS}$ and $CylH_{PKS}$-S1201A were successfully purified from the *E. coli* expression hosts (**Figure 3.2**). The production of $CylH_{PKS}$ and $CylH_{PKS}$-S1201A in *E. coli* required co-expression of chaperones, GroES and GroEL.

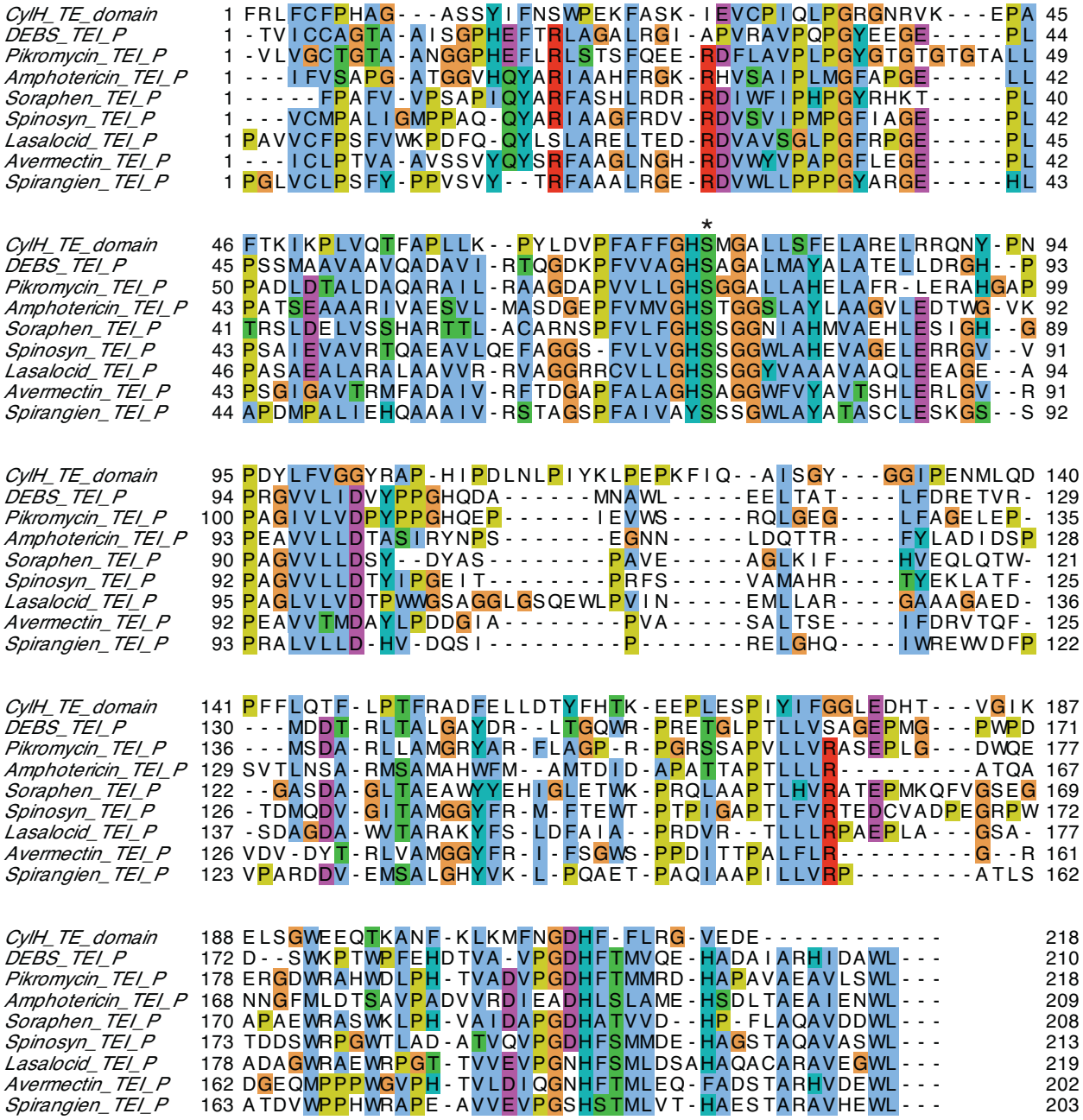


Figure 3.1: Multiple sequence alignment of the CylH TE domain with other previously characterized TE domains of type I PKS enzymes. The catalytic serine residue is marked with asterisk (*).

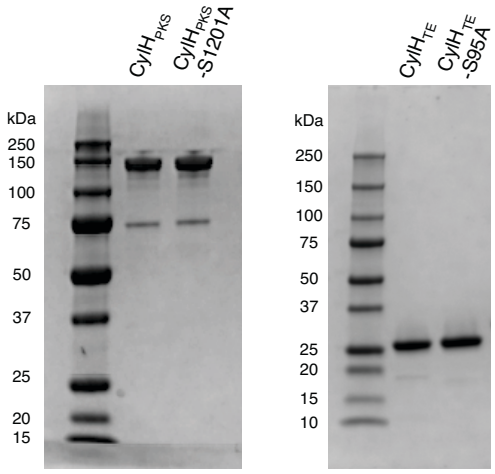


Figure 3.2: SDS-PAGE of purified C-His$_6$-tagged CylH$_{PKS}$ and CylH$_{PKS}$-S1201A (left), as well as CylH$_{TE}$ and CylH$_{TE}$-S95A (right).

Analysis of the purified proteins by gel filtration FPLC showed that CylH$_{PKS}$ exists as a mixture of tetramers and dimers in solution (**Table 3.1** and **Figure 3.3**), while CylH$_{PKS}$-S1201A is primarily a dimer. For wild type CylH$_{PKS}$, the fractions that mainly contained dimers were used for subsequent biochemical assays. Interestingly, CylH$_{TE}$ and CylH$_{TE}$-S95A both exist as monomers in solution, even though TE domains of type I modular PKSs are known to form dimers,[19] On the other hand, freestanding type II TEs are known to exist as monomers,[19] which suggested that the structure of the CylH TE domain might better resemble type II TEs rather than type I TEs.

Table 3.1: Molecular weight analysis of CylH$_{PKS}$, CylH$_{PKS}$-S1201A, CylH$_{TE}$, and CylH$_{TE}$-S95A

Protein	Calculated MW (kDa)	Observed MW (kDa)	Oligomeric state
CylH$_{PKS}$	152.7	694.6 437.1	tetramer dimer
CylH$_{PKS}$-S1201A mutant	152.7	448.1	dimer
CylH$_{TE}$	29.8	27.3	monomer
CylH$_{TE}$-S95A mutant	29.8	25.4	monomer

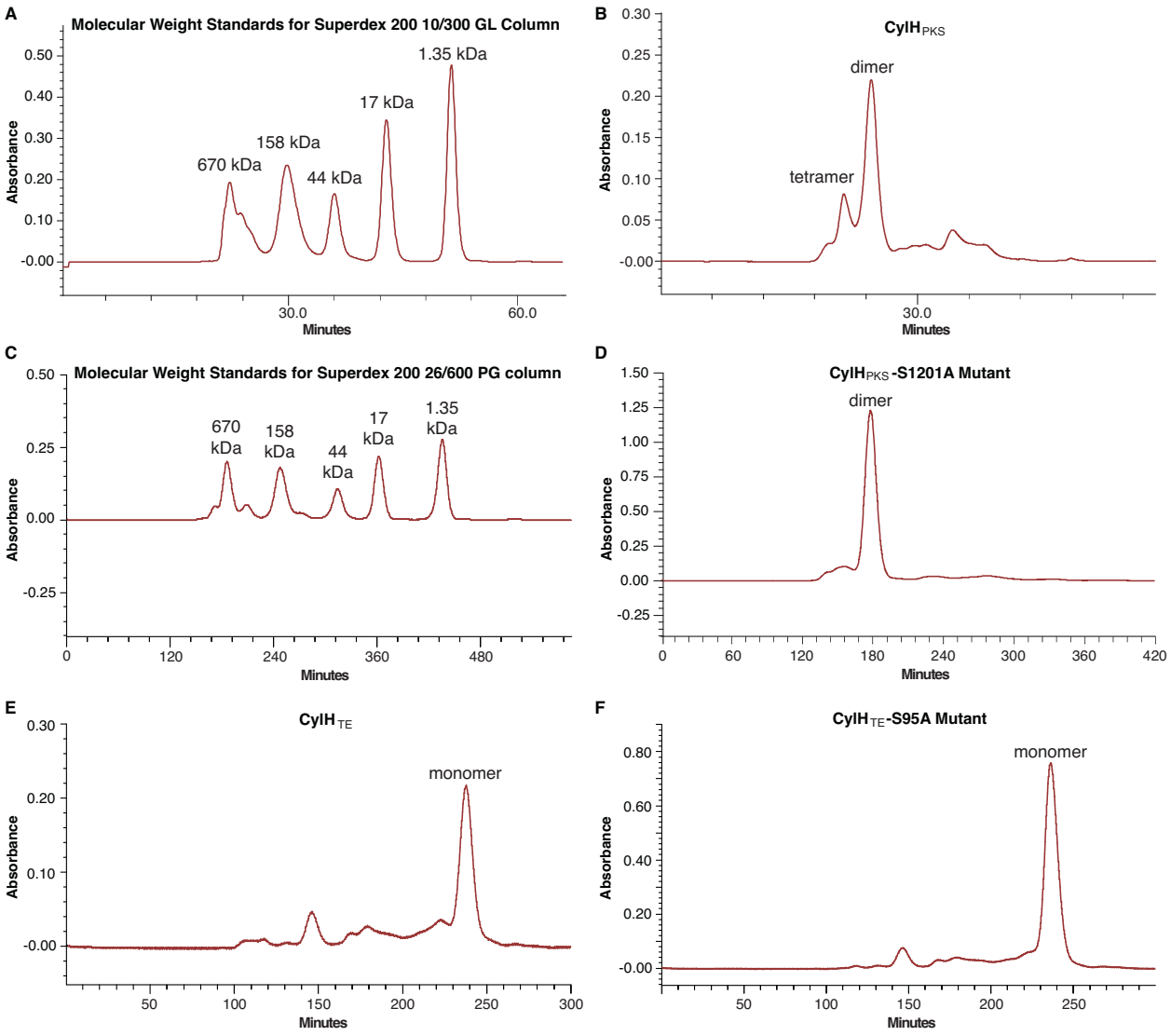


Figure 3.3: Determination of the molecular weights of $CylH_{PKS}$, $CylH_{PKS}$-S1201A, $CylH_{TE}$, and $CylH_{TE}$-S95A by gel filtration FPLC using the Superdex 200 10/300 GL column: **A**) Molecular weight standards, and **B**) $CylH_{PKS}$. FPLC analysis using the Superdex 200 26/600 PG column: **C**) Molecular weight standards, **D**) $CylH_{PKS}$-S1201A mutant, **E**) $CylH_{TE}$, and **F**) $CylH_{TE}$-S95A mutant.

We first verified that the ACP domains of purified $CylH_{PKS}$ and $CylH_{PKS}$-S1201A could be post-translationally modified by Sfp,[30] a promiscuous phosphopantetheinyl transferase, using the BODIPY-CoA (**3.28**) fluorescent phosphopantetheinylation assay (**Figure 3.4A**).[31] The result of this assay indicated that both the wild type $CylH_{PKS}$ and the mutant $CylH_{PKS}$-S1201A could be post-translationally modified by Sfp (**Figure 3.4B**). Relative quantitation of the BODIPY-CoA loading onto $CylH_{PKS}$ and $CylH_{PKS}$-

S1201 showed that wild type CylH$_{PKS}$ was loaded to approximately 80% of that observed for the CylH$_{PKS}$-S1201A mutant (**Table 3.2**).

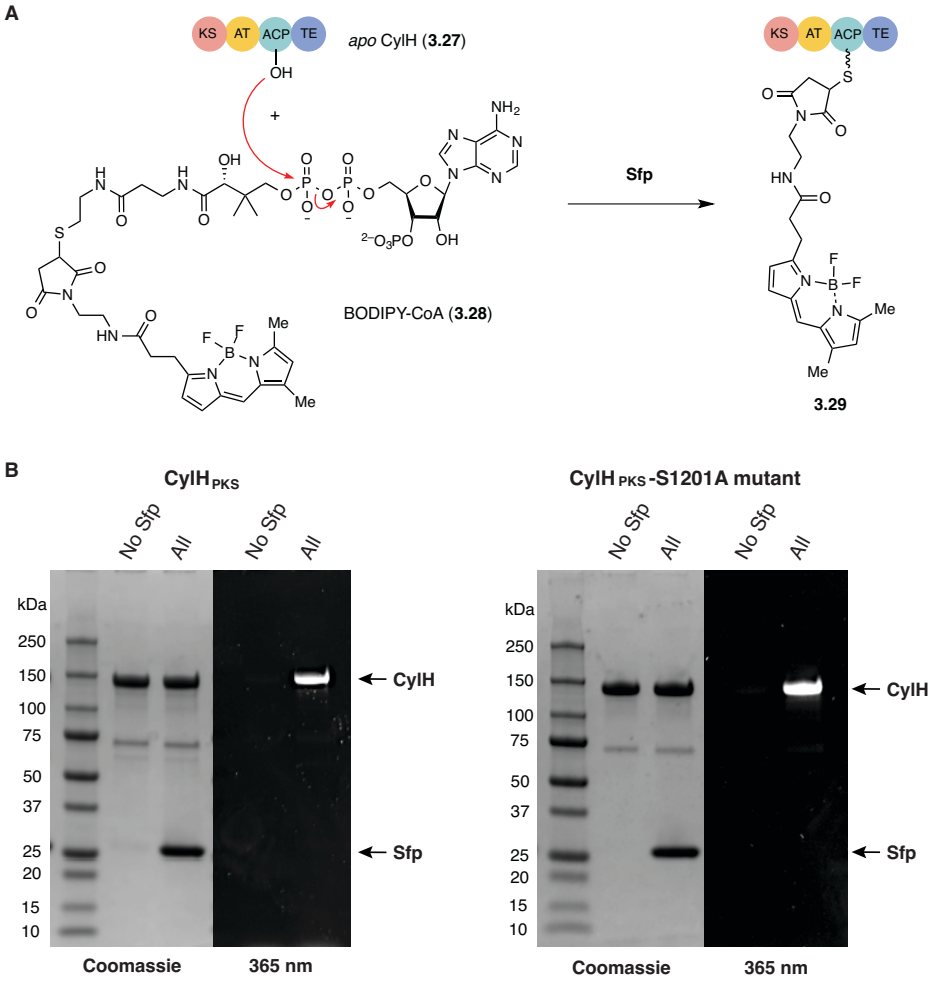
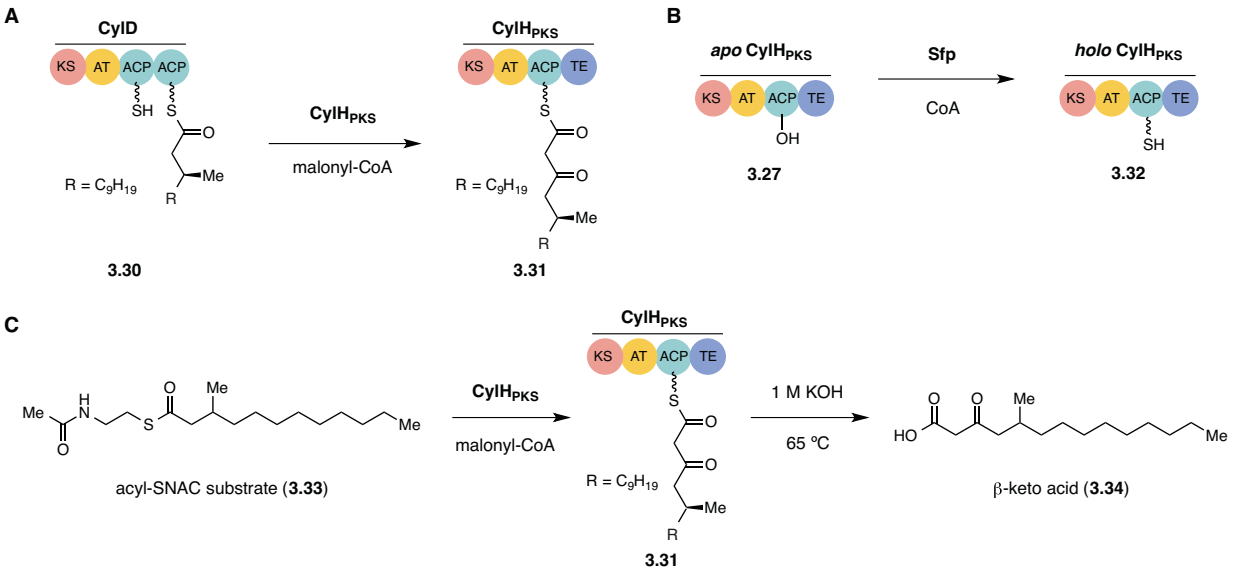


Figure 3.4: BODIPY-CoA fluorescent phosphopantetheinylation assay of CylH$_{PKS}$ and CylH$_{PKS}$-S1201A. **A**) Transfer of the BODIPY-CoA onto the ACP domain of *apo*CylH by PPant transferase Sfp verifies the ability of Sfp to post-translationally modify CylH. **B**) SDS-PAGE of the assay shows that CylH$_{PKS}$ and CylH$_{PKS}$-S1201A are loaded with BODIPY only in the presence of Sfp.

Table 3.2: Relative quantitation of the BODIPY loading assay

Protein	Coomassie signal intensity	BODIPY signal intensity (365 nm)	Signal ratio (BODIPY/Coomassie)	Relative loading
CylH$_{PKS}$	1893474	1916200	1.012	80%
CylH$_{PKS}$-S1201A	2097282	2645045	1.261	100%

To examine the elongation activity of CylH$_{PKS}$, we synthesized acyl-SNAC substrate (**3.33**, see **Scheme 3.8**), which mimics the predicted acyl-ACP substrate (**3.30**) used by CylH. We then performed *in vitro* assays with CylH$_{PKS}$ and acyl-SNAC substrate (**3.33**), analyzing the formation of the elongation product (**3.31**) by LC-MS (**Scheme 3.8C**). The active *holo* forms of CylH$_{PKS}$ (**3.32**) and CylH$_{PKS}$-S1201A were generated by incubating the *apo* enzyme (**3.27**) with CoA and Sfp (**Scheme 3.8B**). The product of the elongation reaction (**3.31**), which is covalently tethered to the ACP domain of CylH$_{PKS}$, was hydrolyzed from CylH$_{PKS}$ by quenching the assay with potassium hydroxide, followed by heating to release the β-keto acid product (**3.34**) for LC-MS analysis (**Scheme 3.8C**). We also prepared a synthetic standard of β-keto acid (**3.34**) for method development on the LC-MS, and confirmed the formation of the β-keto acid product (**3.34**) in the full assay (**Figure 3.5A**). While we anticipated potentially observing spontaneous decarboxylation of β-keto acid (**3.34**) to form the corresponding methyl ketone (**3.35**) during the KOH treatment, we did not detect significant amounts of **3.35** in the assay mixtures (**Figure 3.5B**).



Scheme 3.8: Investigation of the elongation reaction catalyzed by CylH$_{PKS}$. **A**) CylH is predicted to use the acyl-ACP substrate **3.30**, which is generated by the type I PKS assembly line enzymes (CylD-CylG). **B**) The active *holo* form of CylH$_{PKS}$ is formed by incubating the *apo* enzyme with Sfp and CoA. **C**) CylH$_{PKS}$ activity was assayed using the synthetic acyl-SNAC substrate **3.33**. The elongation product is hydrolyzed by treatment with 1 M aq. KOH at 65 °C to release β-keto acid **3.34**.

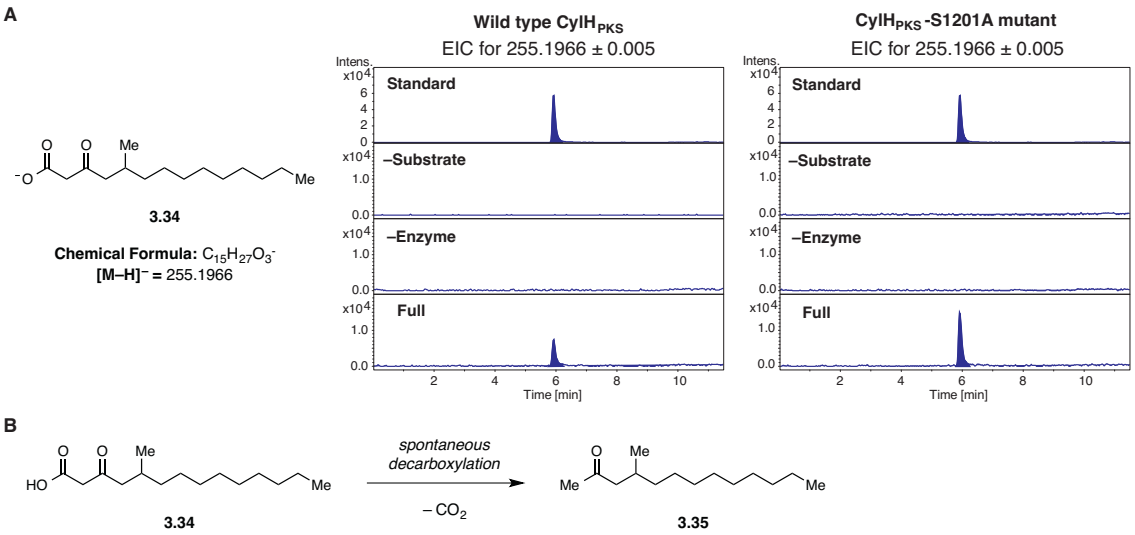


Figure 3.5: LC-MS detection of the CylH elongation product. **A**) β-keto acid (**3.34**) is detected in the full assays for both CylH$_{PKS}$ and CylH$_{PKS}$-S1201A. The peak has the same retention time as the synthetic standard. **B**) The ketone (**3.35**) formed from spontaneous decarboxylation of the β-keto acid (**3.34**) was not consistently detected in the assay samples.

We next examined the activity of CylH$_{PKS}$ together with type III PKS CylI. Addition of CylI to the assay mixture resulted in the production of resorcinol **3.18** (**Figure 3.6**). We observed a small amount of resorcinol formation even in the absence of CylH$_{PKS}$. We propose that this product results from promiscuous activity of CylI toward the acyl-ACP substrate **3.33**, which could form resorcinol **3.18** by catalyzing three extensions with malonyl-CoA prior to cyclization, thioester hydrolysis, decarboxylation and aromatization (**Scheme 3.9**). Nevertheless, the substantial increase in the resorcinol **3.18** formation in the presence of CylH$_{PKS}$ shows that CylI uses the β-ketoacyl-ACP (**3.31**) as a preferred substrate. Overall, these assay results demonstrate that CylH$_{PKS}$ catalyzes elongation using acyl-SNAC substrate (**3.33**) and CylI can utilize a CylH-tethered acyl-ACP substrate (**3.31**).

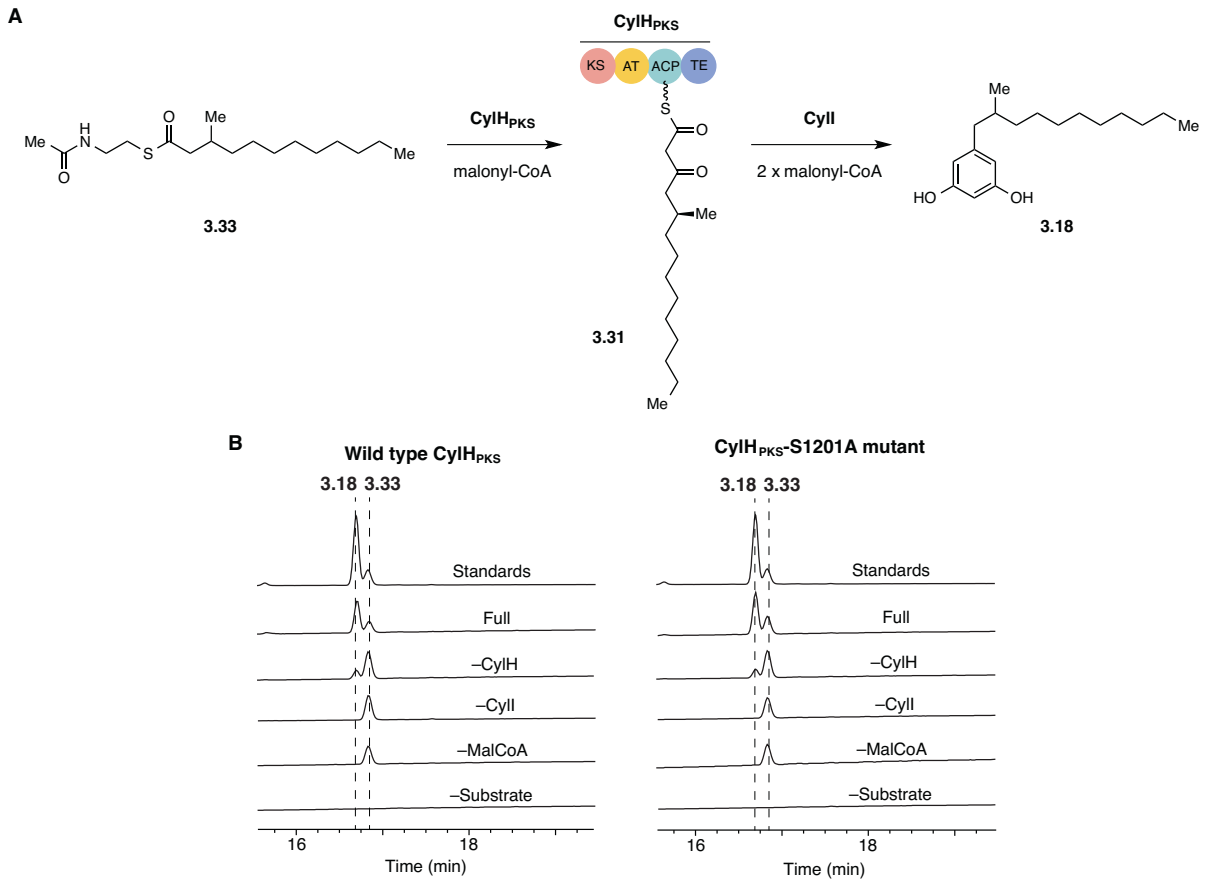
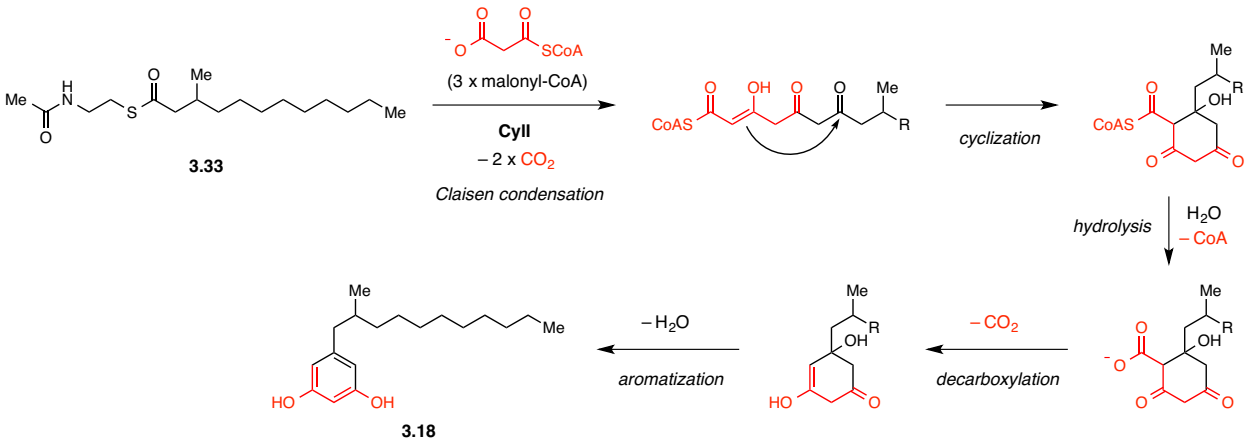


Figure 3.6: Coupled CylH$_{PKS}$ and CylI assay. **A**) CylH$_{PKS}$ forms □-ketoacyl-ACP **3.31** through elongation of the acyl-SNAC substrate **3.33**, and CylI uses **3.31** to form **3.18**. **B**) HPLC analysis of the coupled assays shows that resorcinol formation is observed with both CylH$_{PKS}$ and CylH$_{PKS}$-S1201A mutant. Small amount of resorcinol formation is also seen in the control lacking CylH$_{PKS}$ due to the promiscuity of CylI (monitored at 210 nm).



Scheme 3.9: CylI catalyzes resorcinol formation using the acyl-ACP substrate **3.33** through three elongation cycles, followed by cyclization, decarboxylation, and aromatization.

Qualitative comparison of the CylH and CylI coupled assays performed with $CylH_{PKS}$ and $CylH_{PKS}$-S1201A revealed that the rate of resorcinol formation is not influenced by the presence of the active TE domain (**Figure 3.7**). If the TE domain was catalytically active and had a competing hydrolytic activity toward acyl-ACP **3.31**, then we should have only observed accumulation of the hydrolysis product **3.34** in the assays performed with the wild type $CylH_{PKS}$ with the active TE domain. We therefore quantified the amount of resorcinol **3.18** and the hydrolysis product (**3.34**) in assays containing CylI and either $CylH_{PKS}$ or $CylH_{PKS}$-S1201A by LC-MS (**Figure 3.8A**). The quantitation was performed using the LC-MS QQQ, monitoring the transition specific to **3.18** and **3.34** (**Figure 3.8B**). D_3-resorcinol **3.36** and β-keto acid **3.37** were used as internal standards for the quantitation of resorcinol **3.18** and the hydrolysis product (**3.36**), respectively (**Figure 3.8C**).

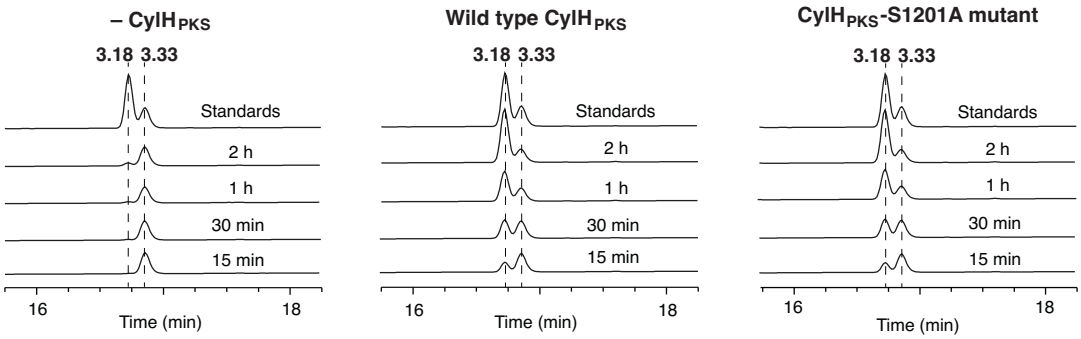


Figure 3.7: Time course analysis of the coupled CylH and CylI assays show that the rate of resorcinol **3.18** formation from **3.33** is qualitatively comparative with $CylH_{PKS}$ and $CylH_{PKS}$-S1201A (monitored at 210 nm).

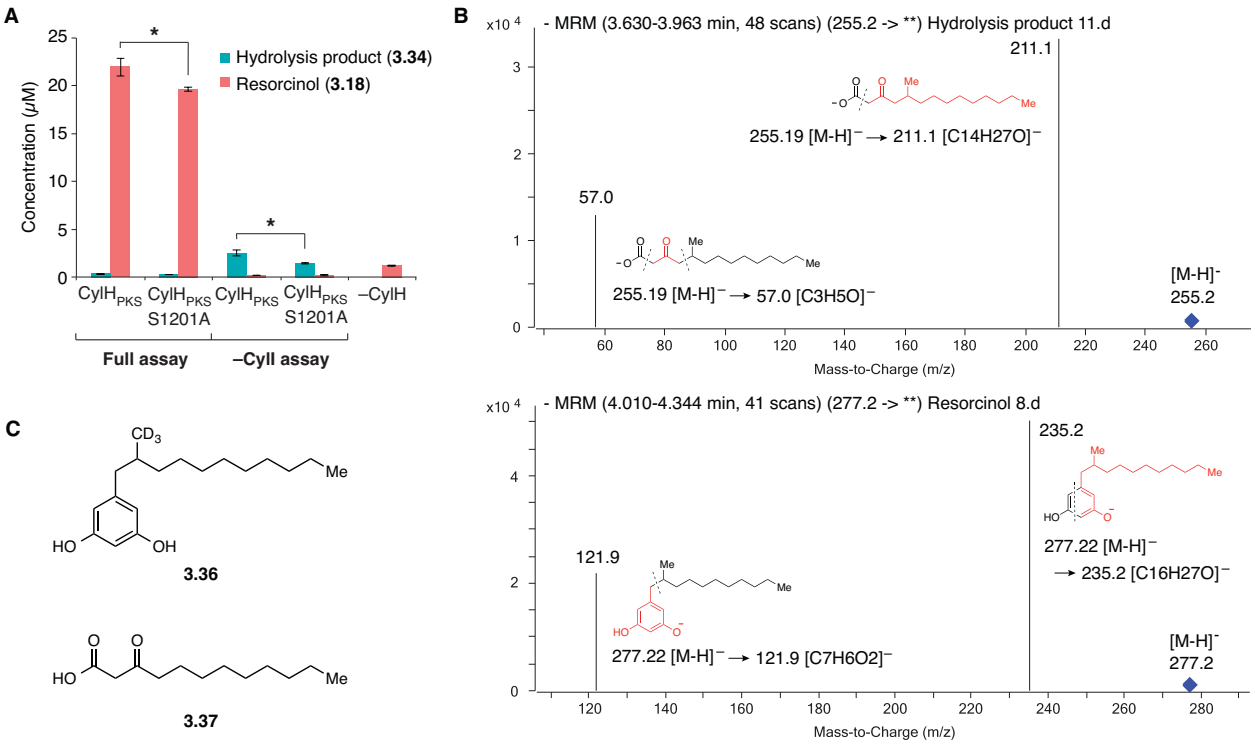
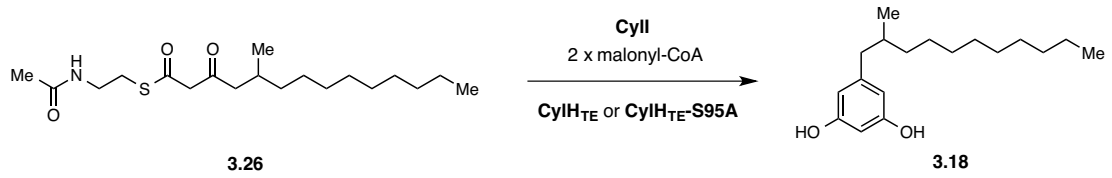


Figure 3.8: Quantitative analysis of the coupled $CylH_{PKS}$ and CylI assays. **A**) Calculated concentrations of the hydrolysis product **3.34** and the resorcinol **3.18** in the assay samples. The quantitation was performed using LC-MS QQQ. **B**) Transitions monitored by LC-MS QQQ for the hydrolysis product **3.34** and the resorcinol **3.18**. **C**) Internal standards used for the quantitation of the hydrolysis product and resorcinol.

In the full assays containing CylI, very small quantities of the hydrolysis product **3.34** were detected (approximately 0.3 μM for both $CylH_{PKS}$ and $CylH_{PKS}$-S1201A) compared to resorcinol **3.18**, indicating that the CylH TE domain does not substantially accelerate the competing hydrolysis of the □-ketoacyl-ACP (**3.31**). The trace amount of hydrolysis product **3.34** observed in these assays could also result from non-enzymatic background hydrolysis of the □-ketoacyl-ACP (**3.31**) if the transfer of the nascent polyketide to CylI is inefficient. In fact, higher concentrations of the hydrolysis product **3.34** are detected in the assay samples lacking CylI, even for the $CylH_{PKS}$-S1201A mutant with an inactive TE domain.

We also explored whether the addition of the excised $CylH_{TE}$ impacted the efficiency of CylI-catalyzed resorcinol **3.18** formation using the □-ketoacyl-SNAC substrate (**3.26**, see **Scheme 3.10**). The addition of $CylH_{TE}$ or $CylH_{TE}$-S95A mutant to the CylI assay did not result in a difference in the qualitative rate of

resorcinol formation (**Figure 3.9**). These results also support the conclusion that the CylH TE domain does not influence product release in the PKS assembly line termination step.



Scheme 3.10: Reaction used to examine the rate of resorcinol formation with and without the excised CylH$_{TE}$.

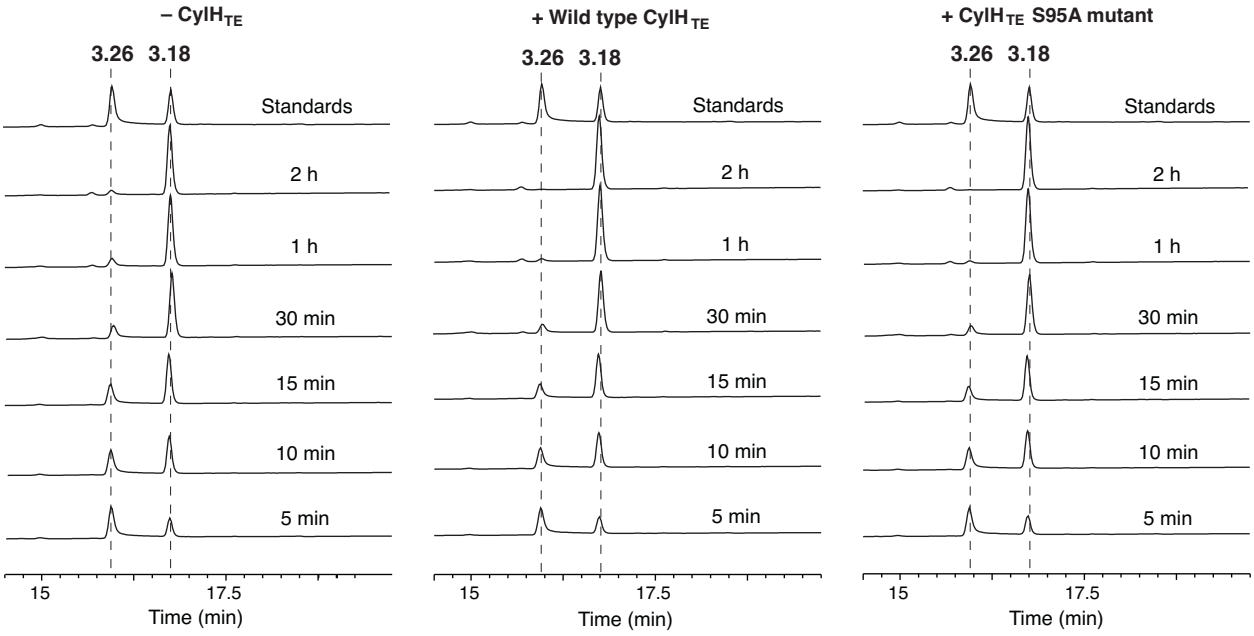


Figure 3.9: Time course analysis of the CylI assay with and without CylH$_{TE}$ or CylH$_{TE}$-S95A (monitored at 230 nm).

3.2.2. Bioinformatics and phylogenetic analysis of the CylH TE domain

To determine whether the apparent lack of CylH TE domain activity arose from mutations to key active site residues, specifically the conserved Ser-His-Asp catalytic triad, we analyzed this domain using bioinformatics. Multiple sequence alignment of the sequence of the CylH TE domain with those of the characterized type I TEs suggested that the catalytic Asp residue might be absent from the CylH TE domain (**Figure 3.10A**). However, when we performed the alignment of the CylH TE domain with previously characterized type II TEs, the catalytic triad of the CylH TE domain appeared to be conserved

(**Figure 3.10B**). This discrepancy between the two alignments again suggested that despite its position at the C-terminus of a type I modular PKS, the CylH TE domain might be more closely related to type II TEs. In addition, the hydrolase signature sequence (GXSXG) of the CylH TE domain matched that of type II TEs perfectly (GHSMG).[32] The hydrolase signature sequence of the CylH TE domain provided additional evidence for its resemblance to type II TEs.

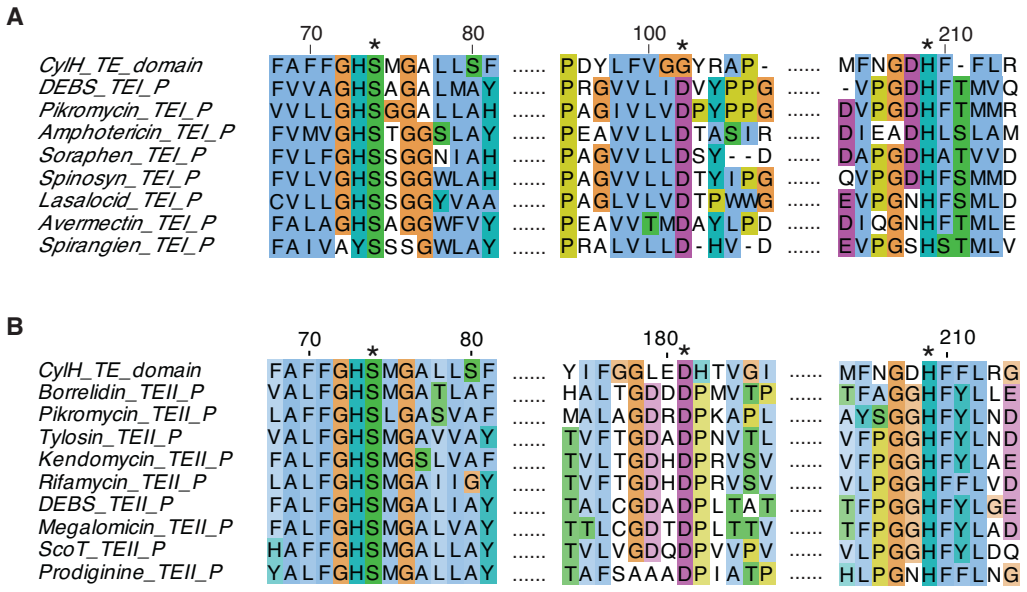


Figure 3.10: Multiple sequence alignments of the CylH TE domain. The catalytic triad (Ser-His-Asp) is marked with asterisk (*) **A**) MSA with type I TEs shows that the catalytic Asp residue appears to be missing. **B**) MSA with type II TEs shows that the catalytic triad is conserved.

Further support for this hypothesis was obtained using an HHPred secondary structural homology search, which identified type II TEs RifR[33] and RedJ[34] as the top two hits. Using the crystal structure of RifR and RedJ as templates, we constructed a homology model of the CylH TE domain. Inspection of this model revealed a well-aligned catalytic triad (**Figure 3.11**).

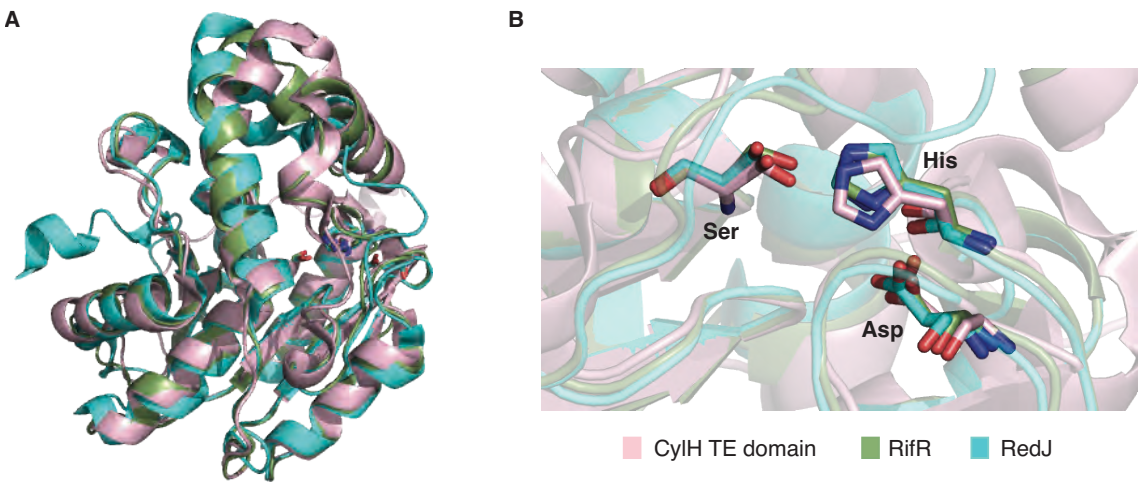


Figure 3.11: Homology model of the CylH TE domain. **A**) Overall structure overlaid with the structures of RifR and RedJ. **B**) The catalytic triad (Ser-His-Asp) is conserved in all three TEs.

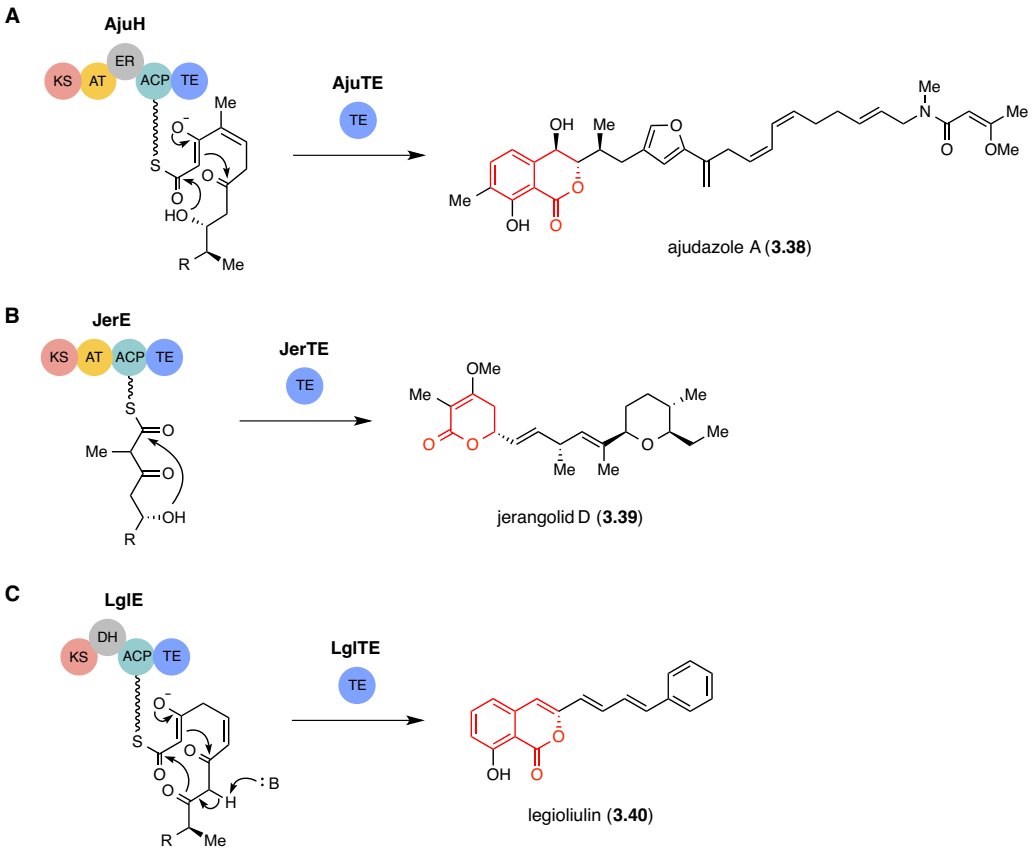
The multiple sequence alignments and the secondary structural homology search result indicated that the CylH TE domain was likely closely related to freestanding type II thioesterases. Similarly, the C-terminal TE domains of the PKS assembly lines involved in ajudazole (AjuTE) and jerangolid (JerTE) biosynthesis have been reported to resemble type II TEs.[32] In fact, the TE domain of CylH appears to have high sequence homology to AjuTE (33% identity, 53% similarity) and JerTE (39% identity, 55% similarity). Unlike the CylH TE domain, however, the hydrolase signature sequence of AjuTE and JerTE differs slightly from those of the type II TEs (GHSMG), and the conserved His residue is replaced by Ser or Cys residue, respectively.[32] In the structural work on RifR, this His residue has been described to stabilize the active site by forming a hydrogen bond with the backbone carbonyl of the catalytic His residue.[33]

Finally, we constructed a phylogenetic tree following the method described by Buntin and coworkers to examine the evolutionary relationship between CylH and previously characterize type I and type II TEs.[32] As we expected, this phylogenetic tree showed that the CylH TE domain clusters with other freestanding type II TEs, and not with type I TEs (**Figure 3.12**). Inclusion of the closest homologs of the CylH TE domain as identified through BLASTp search[35] in this analysis showed that there are other TE domains that are closely related to type II TEs, including AjuTE and JerTE (highlighted in pink in **Figure 3.12**).

Overall, these analyses suggested that the CylH TE domain had the catalytic residues necessary for activity, but was more closely related to type II TEs despite its positioning at the C-terminus of the type I modular PKS.

Figure 3.12: Phylogenetic tree of type I and type II TEs generated using Bayesian estimation reveals clustering of the CylH TE domain with type II TEs. All bootstrap values were greater than 50, and ones with values above 75 are marked with •). TEI = type I TE, TEII = type II TE, P = PKS, N = NRPS, PN = hybrid PKS/NRPS, uk = unknown. TEs that have not been previously characterized are marked with *. Outgroups are the type II TEs from C-1027 and calicheamycin biosynthesis, which are known to significantly differ from other type I and type II TEs.[6, 51]

Previously characterized *in cis* TE domains that have sequence homology to type II TEs (type II TE domains) have been proposed to play a role in assembly line termination. AjuTE, the C-terminal TE domain in ajudazol (**3.38**) biosynthesis, catalyzes isochromanone ring formation (**Scheme 3.11A**).[32] Similarly, homologous C-terminal TE domains in jerangolid (**3.39**)[36] and legioliulin (**3.40**)[37] biosyntheses have been implicated in promoting product release through lactonization (**Scheme 3.11B**) and isochromanone ring formation (**Scheme 3.11C**), respectively. The excised AjuTE and JerTE domains also possess hydrolytic activity toward short chain acyl *p*-nitrophenyl esters (**3.41**-**3.44**, see **Figure 3.13**), substrates commonly used to assess the editing function of type II TEs.[32] More recently, the NRPS assembly line that constructs 4-methyloxazoline-containing natural products was found to contain an unusual module (FmoA5) with two tandem internal TE domains (C-PCP-TE-TE-A-PCP; PCP = peptide carrier protein, equivalent to ACP domain in PKS).[38] The second TE domain has a sequence homology to type II TEs, but the activity of this TE domain has not been characterized.[38] These *in cis* type II TE domains also cluster with type II TEs and the CylH TE domain in our phylogenetic tree.



Scheme 3.11: *In cis* product release catalyzed by type II TE C-terminal domains. **A**) Isochromanone ring formation in ajudazole biosynthesis. **B**) Lactonization in jerangolid biosynthesis. **C**) Isochromanone ring formation in legioliulin biosynthesis.

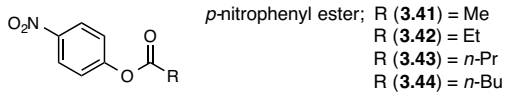


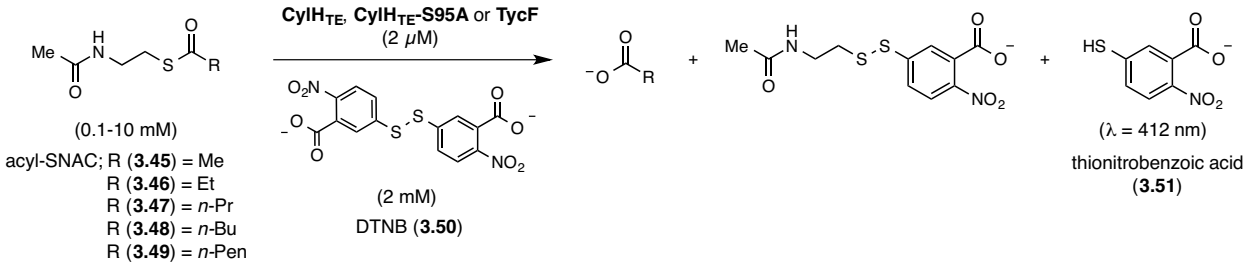
Figure 3.13: Substrates used to screen the hydrolytic activity of AjuTE and JerTE.

3.2.3. Investigation of the editing function of the CylH TE domain

Unlike AjuTE and JerTE, the activity of the CylH TE did not support a role in assembly line termination. Consequently, we investigated the possibility that the CylH TE domain instead had an editing function in the type I modular PKS assembly line. In our LC-MS quantitation of the coupled $CylH_{PKS}$ and CylI assay products, we observed a slight but statistically significant difference in the hydrolytic activities of $CylH_{PKS}$ and $CylH_{PKS}$-S1201A mutant in the absence of CylI (2.5 μM with $CylH_{PKS}$ vs. 1.4 μM with $CylH_{PKS}$-S1201A, see **Figure 3.8A**). This activity appears to be consistent with the functions of type II

editing TEs to reactivate impaired assembly lines by hydrolytic removal of acyl-ACP intermediates with prolonged half-lives.[20,23-25] In addition, CylI produced a slightly greater amount of the resorcinol **3.18** when using the substrate **3.31** generated by the wild type $CylH_{PKS}$ (21.9 μM with $CylH_{PKS}$ vs. 19.5 μM with $CylH_{PKS}$-S1201A). This slight increase in resorcinol formation could result from domain editing activity rescuing a small amount of stalled $CylH_{PKS}$ to maintain efficient polyketide product (**3.31**) formation. As the observed differences were small, we performed additional experiments to investigate the potential editing function of the CylH TE domain.

The editing roles of previously characterized type II TEs have been largely inferred from their ability to hydrolyze short chain acyl-SNAC substrates (**3.45**-**3.49**, see **Scheme 3.12**) that mimic the structures of the acyl-ACP thioesters generated from either premature decarboxylation of extender units or mispriming of *apo* enzymes by promiscuous PPant transferase activity.[25] We therefore examined the activity of $CylH_{TE}$ and $CylH_{TE}$-S95A toward acyl-SNAC substrates (**3.45**-**3.49**) following a previously reported assay procedure (**Scheme 3.12**, **Tables 3.3** and **3.4**).[24] The rate of hydrolysis, which is coupled to the reduction of dithionitrobenzoic acid (DTNB, **3.50**) to thionitrobenzoic acid (**3.51**, absorbance at 412 nm), was monitored using a spectrophotometer. The rate of hydrolysis by $CylH_{TE}$ and $CylH_{TE}$-S95A was compared to that of TycF, a previously characterized type II TE from tyrocidine biosynthesis with known activity toward acetyl-SNAC ($k_{cat}/K_m = 3.6 \pm 0.7\ M^{-1}s^{-1}$, $K_m = 47.9 \pm 9.7$ mM).[24]



Scheme 3.12: Spectrophotometric assay used to analyze the rates of hydrolysis by $CylH_{TE}$, $CylH_{TE}$-S95A and TycF (minotored at 412 nm).

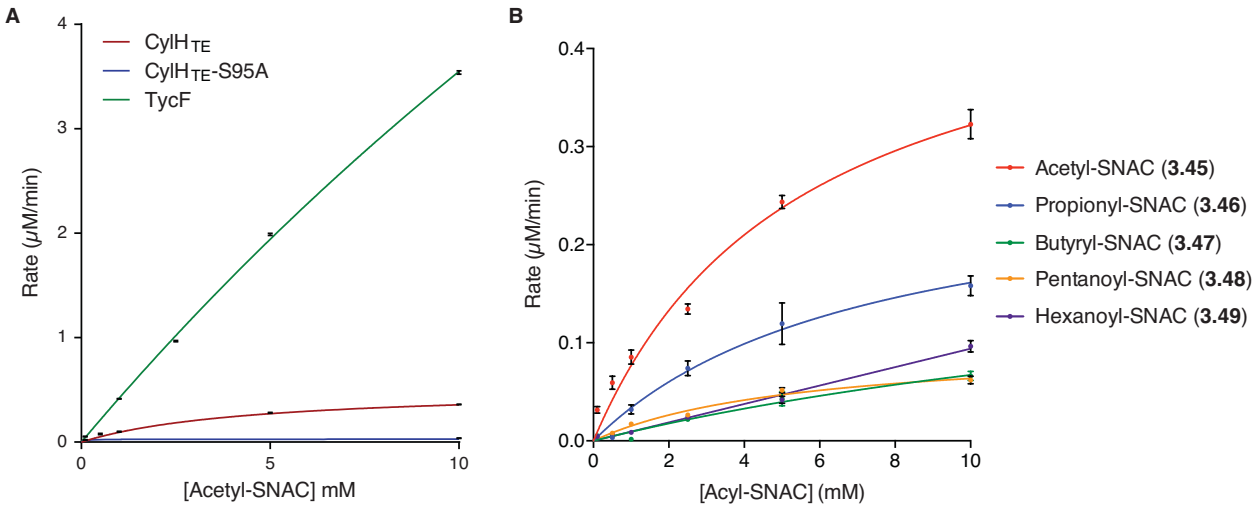


Figure 3.14: Michaelis-Menten curves from the spectrophotometric assays. **A**) Hydrolysis of acetyl-SNAC by CylH$_{TE}$, CylH$_{TE}$-S95A, and TycF performed in triplicate. **B**) Hydrolysis of acyl-SNAC substrates by CylH$_{TE}$ performed in duplicate.

Table 3.3: Kinetic parameters of acetyl-SNAC hydrolysis by CylH$_{TE}$ and TycF (mean ± SD, N = 3).

Enzyme	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (M^{-1}s^{-1})
CylH$_{TE}$	4.5 ± 1.7	0.26 ± 0.04	1.0 ± 0.4
CylH$_{TE}$-S95A	N/A	N/A	N/A
TycF	47.9 ± 9.7	10.28 ± 1.78	3.6 ± 1.0

Table 3.4: Kinetic parameters of CylH$_{TE}$ activity on other short chain acyl-SNAC (mean ± SD, N = 3).

Substrate	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m (M^{-1}s^{-1})
Propionyl-SNAC (**3.47**)	7.3 ± 2.5	0.14 ± 0.03	0.32 ± 0.12
Butyryl-SNAC (**3.48**)	23.1 ± 17.9	0.11 ± 0.06	0.08 ± 0.08
Pentanoyl-SNAC (**3.49**)	5.5 ± 1.8	0.05 ± 0.01	0.15 ± 0.05
Hexanoyl-SNAC (**3.50**)	N/A	N/A	0.08 ± 0.01

CylH$_{TE}$ displayed hydrolytic activity toward acetyl-SNAC (**3.45**, k_{cat}/K_m = 1.0 ± 0.4 M^{-1}s^{-1}, K_m = 4.5 ± 1.7 mM), while CylH$_{TE}$-S95A is inactive (**Figure 3.14A**, **Table 3.3**). CylH$_{TE}$ activity toward acetyl-SNAC (**3.45**), however, was considerably lower than that of TycF. In addition, we tested the hydrolytic activity

of $CylH_{TE}$ on other short-chain acyl-SNAC substrates (**3.46**-**3.49**). $CylH_{TE}$ displayed the highest activity toward acetyl-SNAC (**3.45**), followed by propionyl-SNAC (**3.46**, see **Figure 3.14B** and **Table 3.4**). $CylH_{TE}$ was less active toward longer acyl-SNAC substrates (**3.47**-**3.49**). $CylH_{TE}$'s preference for shorter acyl-SNAC substrates is consistent with an editing function that removes an acyl moiety from the acyl-ACP that would result from premature extender unit decarboxylation or mispriming. Together, these spectrophotometric assays confirmed that $CylH_{TE}$ was catalytically active and suggested a role in assembly line editing.

To investigate the editing function of the CylH TE domain in the context of the cylindrocyclophane type I modular PKS assembly line, we tested its ability to rescue artificially stalled $CylH_{PKS}$. We generated ^{14}C-acetyl $CylH_{PKS}$ (**3.52**) and ^{14}C-acetyl $CylH_{PKS}$-S1201A *in situ* by loading [acetyl-1-^{14}C]-CoA (**3.51**) onto the respective *apo* enzymes using Sfp (**Figure 3.15A**). The amount of radiolabel present on each enzyme at various time points was monitored using gel autoradiography. Comparison of the radiolabelling of $CylH_{PKS}$ and $CylH_{PKS}$-S1201A over time revealed a loss of radiolabel on $CylH_{PKS}$ and accumulation of the radiolabel on $CylH_{PKS}$-S1201A (**Figure 3.15B**). This observation is consistent with the CylH TE domain participating in assembly line editing by hydrolyzing the ^{14}C-acetyl $CylH_{PKS}$ intermediate (**3.52**), similar to the activity of freestanding type II TEs. Overall, the result of the gel autoradiography assay supported the hypothesis that the CylH TE domain plays an editing role in cylindrocyclophane biosynthesis.

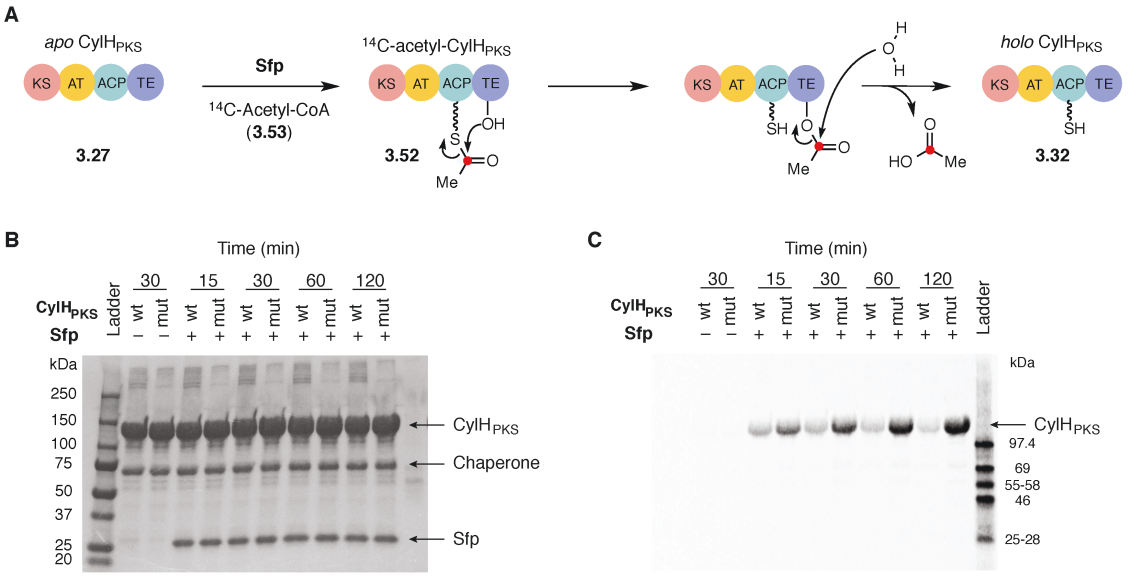


Figure 3.15: Gel autoradiography assay for analyzing the editing function of CylH$_{TE}$. **A**) ^{14}C-acetyl-CylH$_{PKS}$ (**3.52**) is generated by loading ^{14}C-acetyl-CoA (**3.51**) onto *apo* CylH$_{PKS}$ using Sfp. The loss of radiolabel through hydrolysis of the acetyl moiety is monitored. **B**) SDS-PAGE of the gel autoradiography assay visualized with Coomassie stain. **C**) Autoradiogram of the SDS-PAGE shows that the radiolabel accumulates on the mutant CylH$_{PKS}$-S1201A, but the label does not accumulate on the wild type CylH$_{PKS}$.

Overall, our investigations of the CylH TE domain demonstrated that this domain has an editing activity and lacks a role in product release during assembly line termination. A potential explanation for the editing role of the CylH TE domain could be that the activity of the CylH TE domain has evolved to maintain a higher proportion of the *holo* enzyme (**3.32**) in the type I modular PKS assembly line. As the formation of the monomeric resorcinol **3.18** by the type I PKS assembly line appears to be tightly coupled to the subsequent dimerization event,[26] we hypothesize that the editing function of the CylH TE domain may be important for controlling the rate of monomer generation. Furthermore, this assembly line arrangement is intriguing from an evolutionary perspective. The CylH TE domain may have evolved through the fusion of a freestanding type II TE to the C-terminus of the type I modular PKS. Alternately, this domain could represent a snapshot in the evolution of a non-canonical *in trans* termination mechanism, with the presence of the C-terminal TE domain a relic from an ancestral terminating PKS module.

3.2.4. Exploring the distribution of *in cis* type II TE domains in PKS and NRPS assembly lines

Intrigued by the unusual role of the CylH TE domain, we performed additional bioinformatic analyses to determine the prevalence of similar *in cis* type II TE domains in type I PKS or NRPS assembly lines. A BLASTp search[35] of sequenced bacterial genomes using the CylH TE domain as a query led to the identification of 84 unique type I PKS and NRPS assembly line enzymes containing *in cis* type II TE domains, including AjuTE and JerTE (**Table 3.5**). Many of the hits have unusual domains or domain organizations, such as a hybrid PKS/NRPS from *Burkholderia mallei*, which contains TE domains at both its N- and C-termini. The ubiquity of the *in cis* type II TE domains, as well as the diverse activities of those that have been biochemically characterized,[32,36,37] suggests that these domains might possess unexplored functions and could be important new tools for combinatorial biosynthesis. These type II TE domains could also provide editing activities at particular steps in biosynthesis that are important for maintaining the efficiency of these assembly line enzymes.

Table 3.5: BLASTp search result of the CylH TE domain homologs.

#	Organism	Accession number	Length (aa)	Annotated function	Domain organization	E value
1	*Cylindrospermum stagnale*	WP_015207397.1	2241	PKS	ECH-ER-KS-AT-T-TE	1.40E-163
2	*Pleurocapsa* sp. PCC 7319	WP_019504305.1	1593	PKS	KS-AT-DH-T-TE	1.87E-77
3	*Cyanothece* sp. PCC 7424	WP_015954747.1	1337	PKS	KS-AT-T-TE	4.42E-76
4	*Gloeocapsa* sp. PCC 73106	WP_006530113.1	1316	PKS	KS-AT-T-TE	1.14E-74
5	*Chlorogloeopsis fritschii*	WP_016872856.1	1281	PKS	KS-AT-T-TE	4.46E-73
6	*Microcystis aeruginosa*	WP_002776857.1	1320	PKS	KS-AT-T-TE	9.33E-61
7	*Cylindrospermum stagnale*	WP_015328279.1	2172	PKS	KS-AT-MT-KR-T-TE	1.47E-51
8	*Kamptonema* (multispecies)	WP_007358014.1	1312	PKS	KS-AT-T-TE	1.79E-49
9	*Burkholderia glumae*	WP_012732747.1	2381	PKS	KS-AT-DH-EH-KR-T-TE	6.15E-48
10	*Paenibacillus alvei*	WP_005550795.1	1815	NRPS	C-A-T-E-TE	9.76E-48
11	*Stigmatella aurantiaca*	WP_013375988.1	2092	PKS	KS-AT-DH-KR-T-TE	4.43E-44

(Continued)

12	*Chondromyces crocatus*	CAQ18835.1	1620	AjuH (PKS)[32]	KS-AT-DH-T-TE	3.81E-43
13	*Mesorhizobium amorphae*	WP_006199852.1	1488	PKS	TE-A-T-KS-AT	1.00E-42
14	*Gloeobacter violaceus*	WP_011141939.1	963	PKS	KS-T-TE	1.61E-41
15	*Sorangium cellulosum*	ABK32291.1	2869	JerE (PKS)[18]	KS-AT-KR-T-KS-AT-T-TE	2.92E-41
16	*Trichodesmium erythraeum*	WP_011613196.1	1354	PKS	KS-AT-T-TE	3.85E-41
17	*Mesorhizobium* sp. L48C026A00	WP_023803946.1	1920	PKS	TE-A-T-KS-AT-T	6.61E-41
18	*Sorangium cellulosum*	CCE88378.1	2453	PKS	KS-AT-DH-ER-KR-T-TE	3.49E-40
19	*Candidatus Magnetoglobus multicellularis str. Araruama*	ETR71534.1	1348	PKS	KS-DH-T-TE	7.53E-38
20	*Clostridium pasteurianum*	WP_023973746.1	1502	PKS	KS-AT-DH-T-TE	3.31E-37
21	*Rhizobium sullae*	WP_027513632.1	1773	PKS	KS-AT-KR-T-TE	1.06E-36
22	*Pseudomonas corrugata*	WP_024779985.1	1787	PKS	KS-AT-KR-T-TE	5.00E-36
23	*Chondromyces apiculatus* DSM 436	EYF05476.1	4240	PKS	KS-AT-T-KS-AT-KR-T-KS-AT-DH-T-TE	2.88E-35
24	*Sinorhizobium arboris*	WP_028001517.1	1768	PKS	KS-AT-KR-T-TE	3.20E-35
25	*Cyanothece* sp. ATCC 51142	WP_012362254.1	1285	PKS	KS-AT-T-TE	1.42E-34
26	*Bacillus cereus*	WP_002107075.1	1531	PKS	KS-AT-DH-T-TE	3.19E-34
27	*Clostridium* sp. CAG:411	WP_022438322.1	2203	PKS-NRPS	TE-A-KR-T-SDR	2.83E-33
28	*Legionella cherrii*	WP_028380815.1	1293	PKS	KS-DH-T-TE	5.58E-33
29	*Streptomyces longisporoflavus*	ACR50795.1	1502	PKS	KS-AT-DH-T-TE	1.23E-31
30	*Mesorhizobium ciceri*	WP_027038666.1	1758	PKS	KS-AT-KR-T-TE	9.68E-31
31	*Streptomyces mobaraensis*	WP_004946356.1	2674	PKS/NRPS	TE-A-T-KS-AT-TE-KR-T	1.60E-30
32	*Magnetospirillum* sp. SO-1	WP_008618542.1	2056	PKS	KS-AT-DH-KR-T-TE	3.58E-30
33	*Corallococcus coralloides*	WP_014397901.1	1359	PKS	DH-ER-KR-T-TE	6.46E-30
34	*Burkholderia* sp. UYPR1.413	WP_028371022.1	2062	PKS	KS-AT-DH-KR-T-TE	6.93E-30
35	*Nocardia* sp. BMG51109	WP_024805020.1	2730	PKS	KS-AT-DH-MT-ER-KR-T-TE	1.02E-29

(Continued)

36	*Bradyrhizobium japonicum*	WP_028153765.1	2060	PKS	KS-AT-DH-KR-T-TE	1.49E-29
37	*Streptomyces* sp. CNQ329	WP_027770067.1	1499	PKS	KS-DH-T-TE	1.76E-28
38	*Mesorhizobium loti*	WP_010915895.1	1780	PKS	KS-AT-KR-T-TE	3.99E-28
39	*Branchiostoma floridae*	XP_002598380.1	3473	PKS/NRPS	KS-AT-DH-ER-KR-T-TE-C-A	6.23E-28
40	*Pseudomonas aeruginosa*	WP_023124984.1	1392	NRPS	A-MT-T-C-T-TE	1.02E-27
41	*Streptomyces* sp. NRRL 11266	BAE93739.1	1508	PKS	KS-DH-T-TE	1.43E-27
42	*Burkholderia pseudomallei*	WP_024430902.1	2832	PKS/NRPS	TE-A-T-KS-AT-TE-KR-T	2.75E-27
43	*Burkholderia mallei*	WP_004195624.1	2366	PKS/NRPS	TE-A-T-KS-AT-TE	2.77E-27
44	*Paenibacillus* sp. OSY-SE	WP_019419387.1	1660	PKS	KS-AT-DH-T-TE	6.47E-27
45	*Azospirillum lipoferum*	WP_012978027.1	2157	PKS	KS-AT-DH-KS-T-TE	9.99E-27
46	*Fulvivirga imtechensis*	WP_009578358.1	1986	PKS	KS-AT-DH-KR-T-TE	2.23E-26
47	*Ornithinibacillus scapharcae*	WP_010094403.1	1319	PKS	KS-AT-DH-T-TE	3.19E-26
48	*Yersinia rohdei*	WP_004716672.1	1966	PKS	KS-AT-DH-ER-KR-T-TE	4.03E-26
49	*Bradyrhizobium japonicum* SEMIA 5079	AHY53341.1	2299	PKS	KS-AT-DH-KR-T-T-TE	4.31E-26
50	*Mesorhizobium* sp. WSM3224	WP_027170255.1	1843	PKS/NRPS	TE-A-T-KS-AT-T	5.53E-26
51	*Bacillus amyloliquefaciens* SQR9	AHZ14674.1	1632	PKS	KS-AT-DH-T-TE	1.07E-25
52	*Burkholderia thailandensis*	WP_011401503.1	2792	PKS/NRPS	TE-A-T-KS-AT-TE-KR-T	1.30E-25
53	*Nocardiopsis chromatogenes*	WP_026123546.1	2674	PKS	TE-A-T-KS-AT-TE-KR-T	1.47E-25
54	Uncultured bacterium psy1	ADA82585.1	12645	PKS/NRPS	KS-MT-T-C-A-T-KS-KR-T-KS-KR-MT-T-KS-DH-KR-T-KS-KR-MT-T-KS-KR-T-KS-MT-T-KS-T-KS-DH-T-TE	1.75E-25
55	*Branchiostoma floridae*	XP_002610053.1	3311	PKS	KS-AT-DH-KR-T-TE-C	2.62E-25
56	*Burkholderia oklahomensis*	WP_010117951.1	1596	PKS/NRPS	TE-A-T-KS-AT	4.63E-25
57	*Branchiostoma floridae*	XP_002613499.1	1705	PKS/NRPS	KR-T-TE-C-A	6.50E-25

(Continued)

58	*Branchiostoma floridae*	XP_002598386.1	3458	PKS/NRPS	KS-AT-DH-ER-KR-T-TE-C-A	9.03E-25
59	*Chondromyces apiculatus* DSM 436	EYF04558.1	1153	PKS	KS-DH-T-TE	5.47E-24
60	*Burkholderia thailandensis*	WP_009908075.1	1559	PKS/NRPS	TE-A-T-KS-AT	9.36E-23
61	*Hahella ganghwensis*	WP_020406515.1	2382	PKS	FAAL-KS-AT-KR-T-TE	9.67E-23
62	*Coccomyxa subellipsoidea* C-169	XP_005650993.1	15797	PKS	KS-MT-MT-MT-MT-T-T-T-T-T-KS-KR-T-T-KS-DH-ER-KR-T-T-KS-KR-T-KS-KR-T-T-T-KS-KR-T-T-KS-KR-T-KS-DH-KR-T-T-KS-DH-KR-T-T-KS-KR-T-TE-T-TE	1.73E-22 2.03E-22
63	*Propionibacterium propionicum*	WP_014846098.1	1536	PKS	KS-AT-DH-T-TE	1.67E-21
64	*Vibrio caribbenthicus*	WP_009601858.1	1072	NRPS	A-T-TE	2.16E-21
65	*Yersinia kristensenii*	WP_004388876.1	2392	PKS	KS-AT-DH-ER-KR-T-TE	3.35E-21
66	*Streptomyces* sp. A7248	AFS33452.1	1641	SiaI (PKS)[39]	KS-AT-DH-T-TE	1.06E-20
67	*Trichophyton interdigitale* MR816	KDB25511.1	2403	PKS	KS-AT-DH-ER-KR-T-TE	3.86E-20
68	*Mycobacterium tuberculosis* TKK_03_0022	KCD30187.1	881	NRPS	A-T-TE	6.35E-20
69	*Mycobacterium tuberculosis* MAL010108	KBG75468.1	1414	Phenyloxazoline synthase MbtB (NRPS)	T-Cy-A-T-TE	7.03E-20
70	*Mycobacterium bovis*	WP_024458669.1	991	NRPS	A-T-TE	1.06E-19
71	*Mycobacterium tuberculosis*	WP_003901388.1	1280	NRPS	C-A-T-TE	1.53E-19
72	*Mycobacterium bovis*	WP_019283915.1	1414	NRPS	C-A-T-TE	2.02E-19
73	*Arthroderma gypseum* CBS 118893	XP_003175627.1	2406	PKS	KS-AT-DH-ER-KR-T-TE	2.33E-19
74	*Mycobacterium abscessus*	WP_025239395.1	1257	PKS	TE-KS-AT-T	1.07E-18
75	*Streptomyces fungicidicus*	ABD65958.1	8986	Enduracidin biosynthesis EndC (NRPS)[40]	C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-TE-TE	3.45E-18
76	*Saccoglossus kowalevskii*	XP_006823210.1	2198	PKS	AT-DH-ER-KR-T-TE	7.60E-18

(Continued)

77	*Segniliparus rotundus*	WP_013139326.1	1274	PKS	TE-KS-AT-T	2.48E-17
78	*Emiliania huxleyi* CCMP1516	XP_005790240.1	2677	PKS/NRPS	TE-A-T-KS-AT-TE-KR-T	1.88E-14
79	*Fusarium oxysporum* FOSC 3-a	EWY85877.1	2372	PKS	KS-AT-DH-ER-KR-T-TE	2.34E-14
80	*Strongylocentrotus purpuratus*	NP_001239013.1	2606	PKS	KS-AT-DH-ER-KR-T-TE	4.20E-14
81	*Phytophthora parasitica*	ETL32105.1	1311	NRPS	TE-A-T-Re	6.40E-14
82	*Nannochloropsis gaditana*	EWM26301.1	2887	PKS	KS-AT-DH-ER-KR-T-TE	6.02E-13
83	*Phytophthora infestans* T30-4	XP_002905400.1	1311	NRPS	TE-A-T-Re	6.73E-12
84	*Sorangium cellulosum*	ABK32263.1	1773	AmbH (PKS)[36]	KS-AT-KR-T-TE	1.28E-10

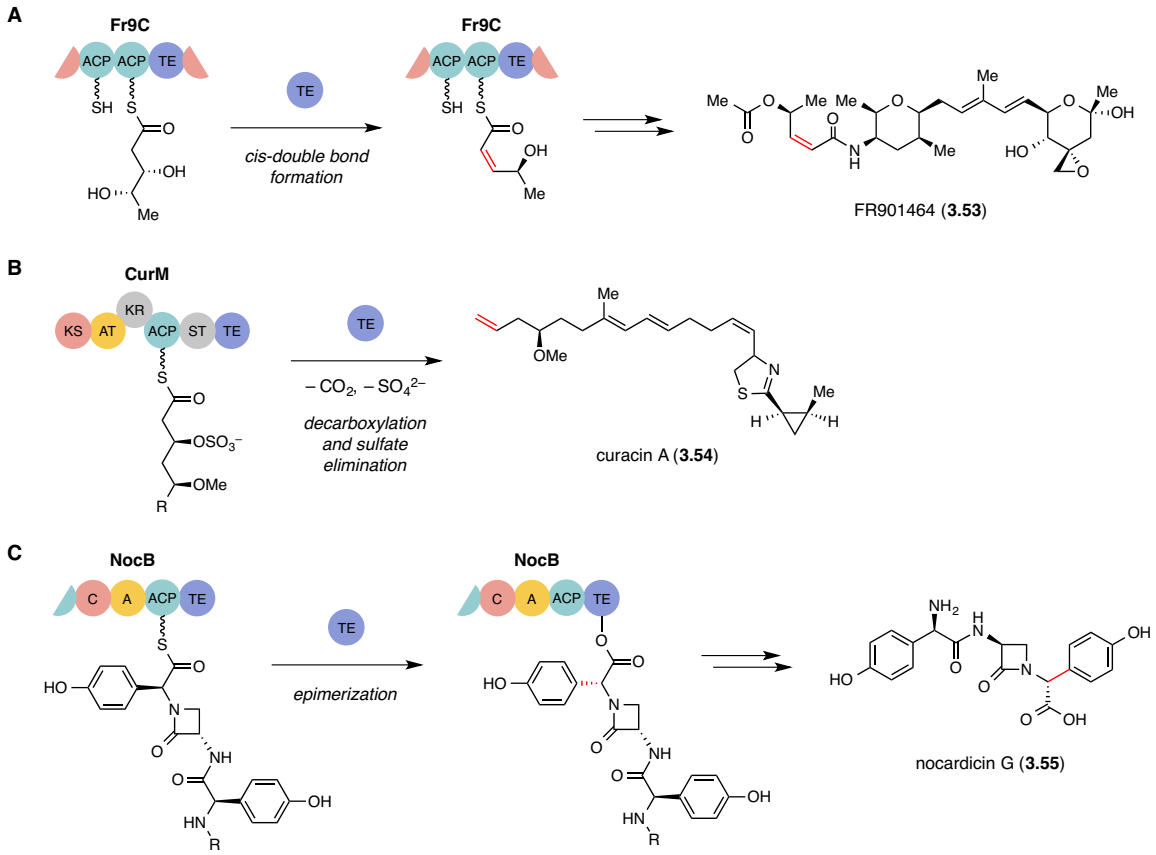
A better understanding of the functions of these non-canonical TE domains could be beneficial for improving the efficiency of engineered assembly lines through strategic introduction of editing TE domains. Finally, the wide distribution of C-terminal type II TE domains has implications for future characterization and annotation of cryptic multimodular assembly line pathways in bacterial genomes, as one can no longer assume that these domains follow canonical termination logic.

3.3. Conclusions

Through this work, we have discovered and characterized the role of the C-terminal *in cis* type II TE domain in the cylindrocyclophane biosynthesis. We have shown that the activity of the CylH TE domain does not influence the assembly line termination step catalyzed by the type III PKS CylI. The CylH TE domain, however, is catalytically active and can hydrolyze short-chain acyl-SNAC substrates as well as acetyl-ACP, which is consistent with this domain having an editing role that preserves assembly line function. The activity of freestanding type II TEs are often important for maintaining the efficiency of PKS and NRPS assembly lines. Likewise, the CylH TE domain might also be necessary for maintaining the productivity of the type I modular PKS in cylindrocyclophane biosynthesis. Similar editing functions are played by *in cis* TE/CLC domain to maintain assembly line fidelity of fungal type I iterative PKS

from aflatoxin (**3.7**, see **Scheme 3.3**) biosynthesis.[41] In addition, the second of the two tandem C-terminal TE domains in the final module of the NRPS assembly line involved in lysobactin biosynthesis (C-A-PCP-TE-TE) has been shown to have an editing function similar to type II TEs.[42] However, unlike the CylH TE domain, the TE/CLC domain in the aflatoxin biosynthesis and the second C-terminal TE domain in the lysobactin biosynthesis have higher sequence homology to type I TEs domains.

Bioinformatic analysis suggests that these atypical *in cis* type II TE domains are widespread in other PKS and NRPS assembly lines. TE domains with sequence homology to type II TEs may play editing functions as in the case of the CylH TE domain, or they may have new unexplored functions. More recently, numerous unusual TE domains that catalyze reactions other than hydrolysis has been discovered.[43-45] For instance, the TE domain in the type I modular PKS assembly line involved in the biosynthesis of FR901464 (**3.55**) has been shown to catalyze a *cis*-double bond formation (**Scheme 3.13A**).[43] The C-terminal TE domain in curacin (**3.56**) biosynthesis is responsible for catalyzing decarboxylation and sulfate elimination along with thioester hydrolysis (**Scheme 3.13B**).[44] Another example is the epimerization catalyzed by the C-terminal TE domain in the biosynthesis of the β-lactam antibiotics (**Scheme 3.13C**).[45] Further efforts to understand the functions of unusual TE domains will improve both our understanding of enzymatic assembly lines and our ability to engineer and utilize the chemical capacities of these systems.



Scheme 3.13: Reactions catalyzed by unusual TE domains. **A**) *Cis*-double bond formation by the TE domain in FR901464 biosynthesis. **B**) Terminal alkene formation by the C-terminal TE domain in curacin biosynthesis. **C**) Epimerization catalyzed by the C-terminal TE domain in nocardicin biosynthesis.

3.4. Materials and Methods

3.4.1. Materials and general methods

All chemicals were obtained from Sigma-Aldrich except for EDC•HCl (Creosalus, Advanced Chemtech), glycerol (Avantor Performance Materials, Macron Fine Chemicals), and [Acetyl-1-^{14}C]-Coenzyme A (Perkin Elmer). Solvents were obtained from Sigma-Aldrich except ethanol (KOPTEC) and diethyl ether (EMD Millipore). Solvents used for organic synthesis (THF, diethyl ether and dichloromethane) were dried and degassed over the solvent column. Luria-Bertani Lennox (LB) medium was purchased from EMD Millipore and IPTG was purchased from Teknova. All NMR solvents were purchased from Cambridge Isotope Laboratories.

Oligonucleotide primers were synthesized by Integrated DNA Technologies (Coralville, IA). Recombinant plasmid DNA was purified with a QIAprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean up were performed using an Illustra GFX PCR DNA and Gel Band Purification kit from GE Healthcare. DNA sequencing was performed by Beckman Coulter Genomics (Danvers, MA). Nickel-nitrilotriacetic acid-agarose (Ni-NTA) resin was purchased from Qiagen and Thermo Scientific. SDS-PAGE gels were purchased from Bio-Rad.

Gel filtration FPLC was performed on a BioLogic DuoFlow Chromatography System (Bio-Rad) equipped with a Superdex column, 200 10/300 GL or 200 26/600 PG (GE Healthcare). Bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa), and vitamin B_{12} (1350 Da) were used as molecular weight markers (Bio-Rad, #151-190). The molecular weights of the proteins analyzed by gel filtration were calculated from their elution volume, using a linear relationship between log(molecular weight) and ve/vo (elution volume/void volume). Protein concentrations were determined using Bradford reagent with bovine serum albumin (BSA) as a standard, or using a NanoDrop 2000 UV-Vis Spectrophotomter (Thermo Scientific) for $CylH_{PKS}$ (ε = 169,820 $M^{-1}cm^{-1}$) and $CylH_{TE}$ (ε = 30,035 $M^{-1}cm^{-1}$). The extinction coefficients for $CylH_{PKS}$ and $CylH_{TE}$ were calculated using ExPASy ProtParam tool (http://www.expasy.org/tools/protparam.html). Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm.

Analytical HPLC was performed on a Dionex Ultimate 3000 instrument (Thermo Scientific) using a Chromolith RP-18e endcapped monolithic silica column (4.6 x 100 mm, EMD Millipore). High-resolution mass spectral data and the spectrophotometric assay data were obtained in the Small Molecule Mass Spectrometry Facility, FAS Division of Science. Mass spectral data were obtained on an Agilent 1290 Infinity UHPLC system (Agilent Technologies) and a Maxis impact UHR time-of-flight mass spectrometer system (Bruker Daltonics Inc) equipped with an electrospray ionization (ESI) source or a 6460 Triple Quadrupole mass spectrometer (Agilent Technologies) equipped with an electrospray

ionization (ESI) source with Agilent Jet Stream technology. Spectrophotometric assay was performed on Spectramax i3 Plate Reader (Molecular Devices).

All synthetic reactions were performed under inert atmosphere. Proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were obtained on a Varian Inova-500 (500 MHz, 125 MHz), Varian-Mercury 400 (400 MHz, 100 MHz) or Varian-Mercury 300 (300 MHz, 75 MHz) NMR spectrometers in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance as internal standard for ^{1}H ($CDCl_3$, δ_H = 7.26 ppm) and ^{13}C ($CDCl_3$, δ_C = 77.2 ppm). Data are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, dd = doublet of doublets, t = triplet, q = quartet, m = multiplet), integration, and coupling constant. NMR spectra were visualized using iNMR Version 5.3.6 (Mestrelab Research).

High-resolution mass spectral (HRMS) data for the synthetic compounds were obtained in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology on a Bruker Micro qTOF-QII fitted with a dual-spray electrospray ionization (ESI) source or in the Small Molecule Mass Spectrometry Facility, FAS Division of Science on an Agilent 6210 TOF. The capillary voltage was set to 4.5 kV and the end plate offset to –500 V, the drying gas temperature was maintained at 190 ºC with a flow rate of 8 L/min and a nebulizer pressure of 21.8 psi. The liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with 50% H_2O and 50% acetonitrile as solvent.

3.4.2. Cloning of CylH$_{PKS}$, CylH$_{PKS}$-S1201A, CylH$_{TE}$, and CylH$_{TE}$-S95A

N-His$_6$-tagged CylI was cloned according to a previously published procedure.[26] pET30a-TycF was obtained from the Walsh group.[24]

Table 3.6: Oligonucleotides used for cloning (restriction sites underlined).

Primers	Sequence	Target
CylH-KS-F-NdeI	5' CCGCCCATATGAATACTGTTATTTCCAAGC 3'	$CylH_{PKS}$ (KS-AT-ACP-TE)
CylH-RC-XhoI-nostop	5' GCCCTCGAGATTGATTGTACAGAAAATAG 3'	
TE-20aa-F-NdeI	5' GCACGACATATGAAGGTTGTAGATGAAGCC 3'	$CylH_{TE}$
TE-RC-XhoI-nostop	5' GTAGGGCTCGAGGAAAATAGTCTCTTCAAT 3'	
CylH-TE-S74A	5' CATTTTTCGGTCATGATATGGGAGCACTACTC 3'	TE active site serine to alanine mutants $CylH_{PKS}$-S1201A and $CylH_{TE}$-S95A
CylH-TE-S74A-RC	5' GAGTAGTGCTCCCATATCATGACCGAAAAATG 3'	

Cloning of *cylH* tetra-domain and TE domain: *CylH*$_{PKS}$ (KS-AT-ACP-TE) and *cylH*$_{TE}$ (TE domain only) were PCR amplified from *Cylindrospermum licheniforme* ATCC 29412 genomic DNA using the primers shown in **Table 3.6**: *cylH*$_{PKS}$ (forward primer **CylH-KS-F-NdeI** + reverse primer **CylH-RC-XhoI-nostop**), and *cylH*$_{TE}$ (forward primer **TE-20aa-F-NdeI** + reverse primer **TE-RC-XhoI-nostop**). All forward primers have *Nde*I restriction sites and all reverse primers have *Xho*I restriction sites. All PCR mixtures contained Phusion High-Fidelity PCR Master Mix (New England BioLabs), 50-100 ng DNA template, and 25 pmoles of each primer in a total volume of 50 µL. Thermocycling was carried out in a C1000 Gradient Cycler (Bio-Rad) using the following parameters: denaturation for 30 sec at 98 ºC, followed by 30 (*cylH*$_{PKS}$) or 35 cycles (*cylH*$_{TE}$) of 10 sec at 98 ºC, 30 sec at 56 ºC, 5 min (*cylH*$_{PKS}$) or 45 sec (*cylH*$_{TE}$) at 72 ºC, and a final extension time of 10 min at 72 ºC. PCR mixtures were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled, and purified.

Amplified fragments were digested with *Nde*I and *Xho*I (New England BioLabs) for 2.5 h at 37 ºC. Digests contained 1 µL water, 3 µL NEB buffer 4 (10x), 3 µL of BSA (10x), 20 µL of PCR product, 1.5 µL of *Nde*I (20 U/µL), and 1.5 µL of *Xho*I (20 U/µL). Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digests were ligated into linearized expression vector using T4 DNA ligase (New England BioLabs). Both *cylH*$_{PKS}$ and *cylH*$_{TE}$ digested inserts were ligated into the pET-29b vector to encode a C-terminal His_6-

tagged construct. Ligations were incubated at 16 ºC overnight and contained 3 µL water, 1 µL T4 Ligase Buffer (10x), 1 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA Ligase (400 U/µL). An aliquot (5 µL) of each ligation reaction was used to transform a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identities of the resulting pET29b-*cylH*$_{PKS}$ and pET29b-*cylH*$_{TE}$ constructs were confirmed by sequencing of purified plasmid DNA. The pET29b-*cylH*$_{PKS}$ and pGro7 (Takara) were co-transformed into chemically competent *E. coli* BL21 (DE3) cells (Invitrogen), while the pET29b-*cylH*$_{TE}$ was transformed into *E. coli* BL21 (DE3) cells by itself. The resulting transformants were stored at –80 ºC as frozen 1:1 LB/glycerol stocks.

Site-directed mutagenesis of the CylH TE domain: Site-directed mutagenesis of the TE domains of CylH$_{PKS}$ and CylH$_{TE}$ were performed using the corresponding oligonucleotides (**CylH-TE-S74A** and **CylH-TE-S74A-RC**) listed in **Table 3.6**. The residue mutated was the catalytic Ser74 residue on the CylH TE domain, which corresponded to Ser1201 on CylH$_{PKS}$ and Ser95 on CylH$_{TE}$. The catalytic serine residue (S74) for mutagenesis was identified through multiple sequence alignment of the CylH TE domain with eight other previously characterized TE domains by ClustalW in Geneious Pro Version 7.1.6 software (see **3.4.11**).[46]

3.4.3. Expression and purification of CylH$_{PKS}$, CylH$_{PKS}$-S1201A, CylH$_{TE}$, and CylH$_{TE}$-S95A

N-His$_6$-tagged CylI was expressed and purified according to a previously published procedure.[26] TycF was expressed and purified following a previously reported procedure.[24]

CylH$_{PKS}$ and CylH$_{PKS}$-S1201A mutant: A 50 mL starter culture of pET29b-*cylH*$_{PKS}$ + pGro7 BL21 *E. coli* was inoculated with frozen stock and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. Overnight cultures were diluted 1:100 into 2 L LB medium containing kanamycin and chloramphenicol. Expressions of GroES and GroEL were induced by addition of 0.5 mg/mL of L-arabinose. The cultures were incubated at 37 ºC with 175 rpm shaking until

OD_{600} = 0.3–0.45. The cultures were then transferred to 15 ºC, protein expression was induced with 100 µM of IPTG when OD_{600} = 0.65–0.75, and the cultures were incubated at 15 ºC for ~20 h.

$CylH_{TE}$ and $CylH_{TE}$-S95A mutant: A 50 mL starter culture of pET29b-*cylH*$_{TE}$ BL21 *E. coli* was inoculated with frozen stock and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL kanamycin. Overnight cultures were diluted 1:100 into 2 L LB medium containing kanamycin. The cultures were incubated at 37 ºC with 175 rpm shaking until OD_{600} = 0.3–0.45, and the cultures were transferred to 15 ºC incubator. The protein expression was induced with 500 µM of IPTG when OD_{600} = 0.65–0.75, and incubated at 15 ºC for ~18 h.

Cells from 2 L of culture were pelleted by centrifugation (6,770 ×*g* for 15 min) and resuspended in 40 mL of lysis buffer ($CylH_{PKS}$ and $CylH_{PKS}$-S1201A mutant = 100 mM potassium phosphate, 300 mM NaCl, pH 8.0; $CylH_{TE}$ and $CylH_{TE}$-S95A = 25 mM Tris•HCl, 500 mM NaCl, 10 mM $MgCl_2$, pH 8.5) supplemented with EDTA-free Pierce Protease Inhibitor Tablets (Thermo Fisher). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 8,000–10,000 psi, and the lysate was clarified by centrifugation (28,960 ×*g* for 30 min). The supernatant was incubated with 2 mL of Ni-NTA resin and 5 mM imidazole for 2 h at 4 ºC. The mixture was centrifuged (3,200 ×*g* for 5 min) and the unbound fraction was discarded. The Ni-NTA was resuspended in 10 mL of wash buffer 1 (lysis buffer containing 20 mM imidazole), loaded into a glass column, and washed with 50 mL of wash buffer 1. $CylH_{PKS}$ and $CylH_{PKS}$-S1201A mutant was washed further with 50 mL of wash buffer 2 (lysis buffer containing 40 mM imidazole). Protein was eluted from the column using a stepwise imidazole gradient in elution buffer (50 mM, 100 mM, 200 mM) and collecting 5 mL fractions. SDS-PAGE analysis (4–15% Mini-PROTEAN TGX precast gel) was employed to determine the presence and purity of protein in each fraction. Fractions containing the desired protein were combined and dialyzed twice against 2 L of dialysis buffer ($CylH_{PKS}$ and $CylH_{PKS}$-S1201A mutant = 50 mM potassium phosphate, 100 mM NaCl, pH 8.0; $CylH_{TE}$ and $CylH_{TE}$-S95A mutant = 25 mM Tris•HCl, 50 mM NaCl, pH 8.5). The collected proteins were further

purified by size-exclusion gel filtration using FPLC. Fractions containing protein were combined and dialyzed against storage buffer ($CylH_{PKS}$ and $CylH_{PKS}$-S1201A mutant = 50 mM potassium phosphate, 100 mM NaCl, pH 8.0; $CylH_{TE}$ and $CylH_{TE}$-S95A mutant = 25 mM Tris•HCl, 50 mM NaCl, 10% glycerol, pH 8.5). The solutions were concentrated using a Spin-X® UF 20 mL centrifugal concentrator with a 100,000 ($CylH_{PKS}$) or 10,000 ($CylH_{TE}$) MWCO membrane (Corning®), and the concentrated protein solutions were frozen with liquid N_2 for storage at –80 ºC. This procedure afforded yields of 1.5–1.7 mg/L for the C-His_6-tagged $CylH_{PKS}$, 1.7–2.5 mg/L of the C-His_6-tagged $CylH_{PKS}$-S1201A mutant, 8.3 mg/L for the C-His_6 tagged $CylH_{TE}$, and 6.0 mg/L for the C-His_6 tagged $CylH_{TE}$-S95A mutant.

Molecular weight analysis of $CylH_{PKS}$, $CylH_{PKS}$-S1201A, $CylH_{TE}$, $CylH_{TE}$-S95A: A ~50 μM solution of purified protein was analyzed by gel filtration during large-scale purification as described in **3.4.1**. The $CylH_{PKS}$ was eluted over 72 min with 50 mM potassium phosphate buffer pH 8, 100 mM NaCl at 0.5 mL/min using Superdex 200 30/100 GL column. $CylH_{PKS}$-S1201A mutant was eluted over 450 min with the same buffer as the wild type $CylH_{PKS}$ at 0.8 mL/min using a Superdex 200 26/600 PG column. $CylH_{TE}$ and $CylH_{TE}$-S95A mutant were eluted over 450 min with 25 mM Tris•HCl pH 8.5, 50 mM NaCl at 0.8 mL/min using a Superdex 200 26/600 PG column. A solution of molecular weight markers was analyzed under the same conditions.

3.4.4. BODIPY-CoA fluorescent phosphopantetheinylation assay

BODIPY-CoA[31] and Sfp[47] were prepared using previously reported procedures. Assay mixture contained 2 μM $CylH_{PKS}$ (wild type or S1201A mutant), 10 μM Sfp, 25 μM BODIPY-CoA, 2 mM $MgCl_2$, 1 mM DTT and 50 mM potassium phosphate pH 8.0 in assay volume of 25 μL. Assays were incubated in the dark at room temperature for 1 h. Assay mixtures were diluted 1:1 in 2x Laemmli sample buffer (Bio-Rad), boiled for 10 min, and then separated by SDS-PAGE on a 4-15% Mini-PROTEAN TGX precast gel. The gel was first imaged at λ = 365 nm, then stained with Coomassie and imaged again. The relative amount of BODIPY-CoA loading for $CylH_{PKS}$ and $CylH_{PKS}$-S1201A was quantified by taking the ratios of the band intensities of Coomassie stain and BODIPY absorption at 365 nm.

3.4.5. LC-MS assay for the elongation activity of $CylH_{PKS}$

A stock solution of *holo* $CylH_{PKS}$ and *holo* $CylH_{PKS}$-S1201A mutant were prepared prior to the assay by incubating 30 μM *apo* $CylH_{PKS}$ or *apo* $CylH_{PKS}$-S1201A, 5 μM Sfp, 250 μM CoA, and 1 mM $MgCl_2$ in 50 mM potassium phosphate buffer pH 8.0 in 100 μL volume for 1 h at room temperature. The elongation assay contained 100 μM acyl-SNAC substrate **3.33**, 24 μM *holo* $CylH_{PKS}$ or *holo* $CylH_{PKS}$-S1201A, and 200 μM malonyl-CoA in 100 mM potassium phosphate buffer pH 8.0 in 100 μL volume. Assay mixtures were incubated at room temperature for 2 h, and the 50 μL of each assay was quenched with 10 μL of 1 M aqueous potassium hydroxide. Quenched samples were heated to 65 ºC for 20 min and then acidified with 10 μL of 1.2 M aqueous hydrochloric acid. 50 μL of cold acetonitrile was added to each sample and the samples were incubated overnight at –20 ºC. Samples were thawed on ice and then centrifuged (16,100 □*g* for 10 min at 4 ºC) before LC-MS analyses using a high accuracy Bruker qTOF mass spectrometer in the Small Molecule Mass Spectrometry Facility, Harvard Faculty of Arts and Sciences (FAS) Division of Science.

The UHPLC equipped with a G4220A binary pump, a built-in vacuum degasser and G4226A high performance autosampler was used for the analysis. 8 μL Samples were analyzed by LC-MS (8 μL injection volume) on an XTerra MS C18 analytical column (2.1 x 50 mm, 3.5 μm) equipped with a guard column (2.1 x 10 mm, 3.5 μm, Waters Corporation) at a flow rate of 0.35 mL/min with the column temperature maintained at room temperature. The following elution conditions were applied: a gradient increasing from 20% to 70% solvent B in solvent A over 5 min, a gradient increasing to 100% solvent B over 0.1 min, 100% solvent B for 5 min, a gradient decreasing to 20% solvent B in solvent A over 0.1 min, and 20% solvent B in solvent A for 5 min. (solvent A = 10 mM ammonium formate and 0.03% ammonium hydroxide in water; solvent B = 0.03% ammonium hydroxide in acetonitrile).

For the MS detection, the ESI mass spectra data were recorded on a negative ionization mode for a mass range of m/z 50 to 1200; calibration mode, HPC; spectra rate, 1.00 Hz; capillary voltage, 3400 V; nebulizer pressure, 30.0 psi; drying gas (N_2) flow, 10.0 L/min; drying gas (N_2) temperature, 220 ºC. A

mass window of ± 0.005 Da was used to extract the ion of $[M-H]^-$. Hydrolysis product **3.34** was detected only in the full assay with mass accuracy within 5 ppm, a match between the observed and theoretical isotopic patterns, and a match between the actual samples and the synthetic standard.

3.4.6. Coupled $CylH_{PKS}$ and CylI HPLC assay

Stock solutions of *holo* $CylH_{PKS}$ and *holo* $CylH_{PKS}$-S1201A were prepared prior to the assay by incubating 50 μM *apo* $CylH_{PKS}$, 5 μM Sfp, 250 μM CoA, and 1 mM $MgCl_2$ in 50 mM potassium phosphate buffer pH 8.0 in 100 μL volume for 1 h at room temperature. A typical CylI assay contained 100 μM substrate **3.33**, 20 μM *holo* $CylH_{PKS}$ or *holo* $CylH_{PKS}$-S1201A, 1.2 μM CylI, 500 μM malonyl-CoA, and 1 mM EDTA in 100 mM potassium phosphate buffer pH 8.0 in a final assay volume of 100 μL. The reaction was incubated at room temperature and 30 μL of the reaction was quenched with 60 μL of cold acetonitrile after 15 min, 30 min, 1 h, and 2 h. The quenched sample was incubated on ice for 10 min and then centrifuged (16,100 □*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC (80 μL injection volume) on a Chromolith RP-18e column (4.6 x 100 mm, EMD Millipore) at a flow rate of 1 mL/min and monitoring at 210 nm. The following elution conditions were applied: 20% solvent B in solvent A for 2 min, a gradient increasing to 100% solvent B in solvent A over 15 min, 100% solvent B for 5 min, a gradient decreasing to 20% solvent B in solvent A over 1 min, 20% solvent B in solvent A for 6 min (solvent A = water; solvent B = acetonitrile).

3.4.7. LC-MS quantitation of resorcinol and hydrolysis product formation

A stock solution of *holo* $CylH_{PKS}$ and *holo* $CylH_{PKS}$-S1201A mutant were prepared prior to the assay by incubating 50 μM *apo* $CylH_{PKS}$, 5 μM Sfp, 250 μM CoA, and 1 mM $MgCl_2$ in 50 mM potassium phosphate buffer pH 8.0 in 100 μL volume for 1 h at room temperature. A typical quantitation assay contained 100 μM substrate **3.35**, 20 μM *holo* $CylH_{PKS}$ (wild type or S1201A mutant), 1.2 μM CylI, 500 μM malonyl-CoA, and 1 mM EDTA in 100 mM potassium phosphate buffer pH 8.0 in a final assay volume of 50 μL. The reaction was incubated at room temperature for 2 h, and then 10 μL of an internal standard stock solution containing 120 μM d_3-resorcinol **3.36** and 120 μM hydrolysis product internal

standard **3.37** was added to each sample. Each assay was quenched with 50 µL of cold acetonitrile, and then the quenched sample was incubated overnight at –20 ºC. The sample was thawed on ice, centrifuged (16,100 □*g* for 10 min at 4 ºC), and analyzed using an Agilent QQQ mass spectrometer in the Small Molecule Mass Spectrometry Facility, Harvard Faculty of Arts and Sciences (FAS) Division of Science.

The UHPLC equipped with a G4220A binary pump, a built-in vacuum degasser and G4226A high performance autosampler was used for the analysis. Calibration standards and samples (4 µL injection volume) were analyzed using an ACQUITY^{TM} UPLC BEH C_{18} analytical column (2.1 x 50 mm, 1.7 µm) equipped with a VanGuard BEH C_{18} Pre-column (2.1 x 5 mm, 1.7 µm, Waters Corporation) at the flow rate of 0.6 mL/min. The column temperature was maintained at room temperature. The following elution conditions were applied: a gradient increasing from 20 to 65% solvent B in solvent A for 2.91 min, a gradient increasing from 65 to 100% solvent B in solvent A over 1.01 min, 100% solvent B for 3.5 min, a gradient decreasing to 20% solvent B in solvent A over 0.1 min, 20% solvent B for 2.35 min. (solvent A = 5 mM ammonium hydroxide in water; solvent B = 5 mM ammonium hydroxide in acetonitrile).

For the MS detection, the ESI mass spectra data were recorded on a negative ionization mode by MRM. The precursor → product ions MRM transition used for hydrolysis product **3.34** were m/z 255.19 $[M-H]^-$ → 211.1 $[C_{14}H_{27}O]^-$ (fragmentor voltage 115 V, collision energy (CE) = 12 V) as the quantifier and m/z 255.19 $[M-H]^-$ → 57 $[C_3H_5O]^-$ (fragmentor voltage 115 V, CE = 24 V) as the qualifier; for the hydrolysis product internal standard **3.37**, they were m/z 213.15 $[M-H]^-$ → 169 $[C_{11}H_{21}O]^-$ (fragmentor voltage 110 V, CE = 8 V) as the quantifier and m/z 213.15 $[M-H]^-$ → 153 $[C_{10}H_{17}O]^-$ (fragmentor voltage 110 V, CE = 16 V) as the qualifier; for resorcinol **3.18** they were m/z 277.22 $[M-H]^-$ → 235.2 $[C_{16}H_{27}O]^-$ (fragmentor voltage 155 V, CE = 20 V) as the quantifier and m/z 277.2 $[M-H]^-$ → 121.9 $[C_7H_6O_2]^-$ (fragmentor voltage 155 V, CE = 24 V) as the qualifier; for d_3-resorcinol **3.36** they were m/z 280.24 $[M-H]^-$ → 238.2 $[C_{16}H_{24}D_3O]^-$ (fragmentor voltage 155 V, CE = 20 V) as the quantifier and m/z 277.2 $[M-H]^-$ → 121.9 $[C_7H_6O_2]^-$ (fragmentor voltage 155 V, CE = 24 V) as the qualifier. Nitrogen was used as the drying, sheath, and collision gas. All the source and analyzer parameters were optimized using Agilent

MassHunter Source and iFunnel Optimizer and Optimizer software, respectively. The source parameters are as follows: drying gas temperature 320 ºC, drying gas flow 13 L/min, nebulizer pressure 45 psi, sheath gas temperature 300 ºC, sheath gas flow 11 L/min, capillary voltage 3000 V, and nozzle voltage 0 V. The UHPLC eluant before 0.5 min was diverted to waste.

Standard curves for the hydrolysis product **3.34** and resorcinol **3.18** were generated by analyzing known concentrations of the product standards using the LC-MS method described above. Each sample was spiked with an internal standard stock solution containing 120 µM d_3-resorcinol **3.36** and 120 µM hydrolysis product internal standard **3.37**. The peak area detected for resorcinol **3.18** and the hydrolysis product **3.34** was normalized using the peak area of the internal standards **3.36** or **3.37**, respectively. The concentration of each product vs. ratio of the peak areas (product/internal standard) was plotted to generate the standard curves. The concentration of resorcinol **3.18** and the hydrolysis product **3.34** in each assay sample was calculated from the ratio of the peak areas (product/internal standard) based on the standard curve.

3.4.8. HPLC assay for CylI-catalyzed resorcinol formation in presence of $CylH_{TE}$

Each assay contained 100 µM substrate **3.26**, 10 µM of $CylH_{TE}$ or $CylH_{TE}$-S95A, 1.2 µM CylI, 250 µM malonyl-CoA, and 1 mM EDTA in 100 mM potassium phosphate buffer pH 8.0 in a final assay volume of 100 µL. The reaction was incubated at room temperature and a 30 µL reaction aliquot was quenched with 60 µL of cold acetonitrile after 5 min, 10 min, 15 min, 30 min, 1 h, and 2 h. The quenched sample was incubated on ice for 10 min and centrifuged (16,100 □*g* 10 min at 4 ºC). Supernatant was analyzed by HPLC (80 µL injection volume) on a Chromolith RP-18e column (4.6 x 100 mm, EMD Millipore) at a flow rate of 1 mL/min. The following elution conditions were applied: 20% solvent B in solvent A for 2 min, a gradient increasing to 100% solvent B in solvent A over 15 min, 100% solvent B for 5 min, a gradient decreasing to 20% solvent B in solvent A over 1 min, 20% solvent B in solvent A for 6 min (solvent A = water; solvent B = acetonitrile).

3.4.9. Spectrophotometric assay for the hydrolytic activity of $CylH_{TE}$

A continuous spectrophotometric assay using dithionitrobenzoic acid (DTNB) reduction to thionitrobenzoic acid was employed to monitor generation of free thiol after thioester hydrolysis. Each assay contained 50 mM tris•HCl buffer pH 8, 50 mM NaCl, 2 mM DTNB, 50 μM EDTA, 2 μM enzyme ($CylH_{TE}$, $CylH_{TE}$-S95A, or TycF), and acyl-SNAC substrates **3.45**-**3.49** (0.1 mM, 0.5 mM, 1 mM, 2.5 mM, 5 mM, or 10 mM) in a final volume of 100 μL. The reaction was initiated by adding acetyl-SNAC and monitored at 412 nm by Spectramax i3 Plate Reader. Michaelis-Menten kinetics were calculated employing the value ε = 14,140 $M^{-1}cm^{-1}$ and plotted using Graphpad Prism Version 6. The signal due to background hydrolysis of acetyl-SNAC was taken into consideration by subtracting the averaged signal of the negative controls lacking an enzyme performed in triplicate.

3.4.10. Gel autoradiography assay for analyzing the editing function of $CylH_{PKS}$

^{14}C-Acetyl-ACP was generated by activation with [acetyl-1-^{14}C]-CoA (**3.51**) using the phosphopantetheinyl transferase Sfp. A typical assay contained 10 μM $CylH_{PKS}$ or $CylH_{PKS}$-S1201A, 8 μM Sfp, 50 μM ^{14}C-acetyl-CoA (6.25 nCi), 1 mM $MgCl_2$, and 1 mM TCEP in 50 mM potassium phosphate buffer pH 8. The assay was initiated by addition of ^{14}C-acetyl-CoA **3.51** and incubated at room temperature. After 15 min, 30 min, 1 h, and 2 h, a 10 μL reaction aliquot was removed and quenched by dilution into 10 μL of 2x SDS-loading buffer (no DTT or β-mercaptoethanol added). The quenched samples were heated at ~90 ºC for 10 min and then 10 μL of each sample was run on a SDS-PAGE gel (4-15% Mini-PROTEAN TGX precast gel). The gel was washed with water, stained with Coomassie, destained, dried using a gel dryer (Labconco), and then visualized by phosphorimaging (GE Typhoon Imager, FAS Center for Systems Biology). Reactions with $CylH_{PKS}$-S1201A or lacking Sfp were used as negative controls. DTT and β-mercaptoethanol reducing agents were excluded from the loading dye since they are known to cleave acyl-ACP species.[48]

3.4.11. Generation of the homology model for CylH TE domain using HHpred

CylH TE domain was used as the query to search for structural homologs using HHpred.[49] The two highest hits were known type II TEs from rifamycin and prodiginine biosynthetic pathways (RifR[33] and RedJ,[34] respectively). A homology model of the CylH TE domain was generated by Modeller[15] using the crystal structure of RifR as a template and the resulting PDB file was aligned with the RifR PDB file (3FLA) using MacPyMOL version 1.7 (Schrödinger, Inc.).

3.4.12. Multiple sequence alignment of the CylH TE domain and construction of the phylogenetic tree

Fifty-two amino acid sequences of type I and type II thioesterases from various PKS, NRPS, and hybrid PKS/NRPS biosynthetic systems were obtained from the NCBI public database (http://www.ncbi.nlm.nih.gov/), and the domain boundaries were determined using PKS/NRPS Analysis Web Server (http://nrps.igs.umaryland.edu/nrps/). Multiple sequence alignment for the phylogenetic tree was performed by MUSCLE alignment using Geneious Pro Version 7.1.6.[46]

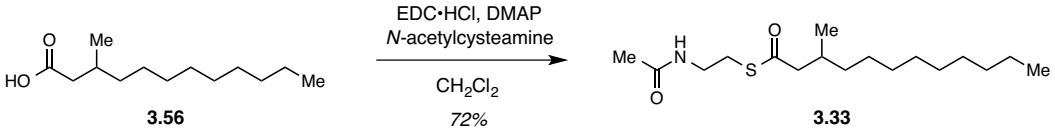
The amino acid sequences were aligned by MUSCLE, and phylogenetic trees were created using MrBayes[50] provided by Geneious Pro Version 7.1.6[46] following a method described in the study of the TE domain in the ajudazole biosynthetic pathway.[32] The Bayesian inference method used the Poisson process with an outgroup set as a type II thioesterase C-1027 from the enediyne polyketide biosynthetic pathway, which has a hotdog fold unlike other TEIIs and TEIs.[51] Markov chain Monte Carlo analysis (MCMC) was performed with 1.1 million generations and four independent chains, and the Markov chain was sampled every 200 generations.

3.4.13. BLASTp search for other TE domains with sequence homology to type II TEs

The CylH TE domain amino acid sequence was used as the query for a BLASTp search on NCBI.[35] Any protein hits with lengths shorter than 300 amino acids were removed to eliminate freestanding type II thioesterases from the search results. The search results were manually curated to remove hits that

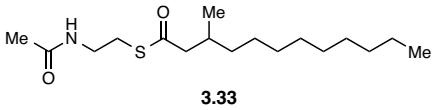
appeared to be identical from the organisms that have multiple protein sequences in the NCBI database. All the hits that did not appear to be type I PKS or NRPS were removed from the search results to give 84 unique hits. The domain organization of each protein hit was determined using the Conserved Domain Database on NCBI[52] and the PKS/NRPS Analysis web server (http://nrps.igs.umaryland.edu/nrps/). Protein hits with an E value greater than 1 E-10 were removed from the list, and the TE domain hits were confirmed to have higher homology to type II TEs by phylogenetic analysis using the Mr. Bayes tool[50] and multiple sequence alignment using the MUSCLE alignment tool on Geneious.[46]

3.4.14. Synthesis of the acyl-SNAC substrate used in CylH assays



Scheme 3.14: Synthesis of the acyl-SNAC substrate **3.33** for CylH assays.

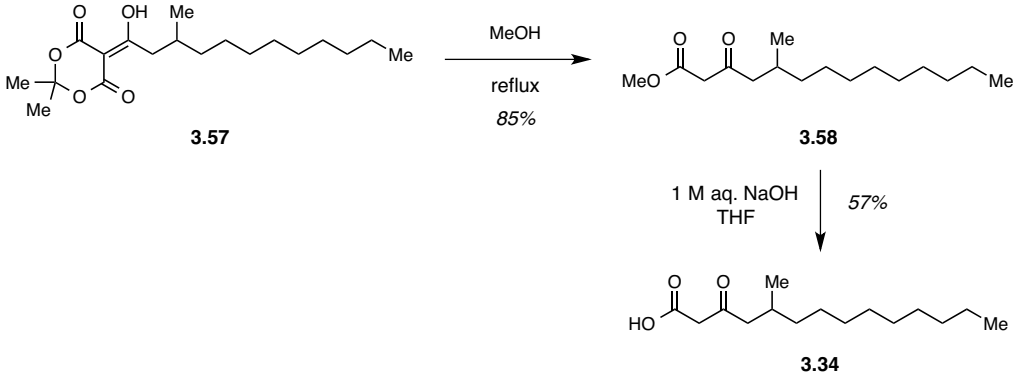
Synthesis of 3-methyldodecanoic acid (**3.56**) is described in **Chapter 2.4.14**.



***S*-(2-acetamidoethyl)-3-methyldodecanethiolate (3.33)**: EDC•HCl (89 mg, 0.47 mmol) was added to a solution of 3-methyldodecanoic acid **3.56** (50 mg, 0.233 mmol) in dichloromethane (2.5 mL), and the resulting mixture was cooled to 0 ºC. The mixture was stirred at 0 ºC for 20 min, and then *N*-acetylcysteamine (0.030 mL, 0.280 mmol) and 4-dimethylaminopyridine (2 mg, 0.016 mmol) were added to the reaction flask. The reaction mixture was warmed to room temperature. After 3 hours, the reaction mixture was quenched with the addition of water (10 mL), and then extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo.* The crude residue was purified with flash chromatography, eluting with 75–100% ethyl acetate in hexanes to afford *S*-(2-acetamidoethyl)-3-methyldodecanethioate **3.33** (53 mg, 0.17 mmol, 72%). TLC: R_f = 0.24 (silica gel, 1:1 hexanes:ethyl acetate). HRMS (ESI): calc'd for $C_{17}H_{34}NO_2S^+$

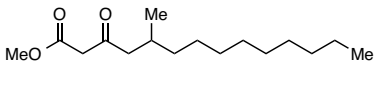
[M+H]$^+$, 316.2305; found, 316.2306. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 5.98 (s, 1H, N**H**), 3.41 (q, 2H, *J* = 6.2 Hz, NHC**H**$_2$), 3.01 (t, 2H, *J* = 6.5 Hz, $CH_2C\mathbf{H}_2S$), 2.55 (dd, 1H, *J* = 14.5 Hz, 5.9, SCOC**H**$_2$), 2.36 (dd, 1H, *J* = 14.5, 8.2 Hz, SCOC**H**$_2$), 1.98 (m, 1H, C**H**), 1.95 (s, 3H, NHCOC**H**$_3$), 1.16-1.29 (m, 16H, C**H**$_2$), 0.91 (d, 3H, *J* = 6.7 Hz, CHC**H**$_3$), 0.86 (t, 3H, *J* = 6.9 Hz, $CH_2C\mathbf{H}_3$). ^{13}C-NMR (125 MHz, CDCl3) δ: 199.8, 170.2, 51.3, 39.8, 36.6, 31.8, 31.1, 29.7, 29.6, 29.5, 29.3, 28.4, 26.8, 23.1, 22.6, 19.4, 14.1.

3.4.15. Synthesis of the hydrolysis product standard



Scheme 3.15: Synthesis of the β-keto acid **3.34** as a product standard for CylH elongation assays.

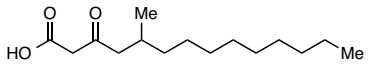
The synthesis of 5-(1-Hydroxy-3-methyldodecylidene)-2,2-dimethyl-1,3-dioxane-4,6-dione (**3.57**) is described in **Chapter 2.4.14**.



3.58

Methyl 5-methyl-3-oxotetradecanoate (3.58): 5-(1-Hydroxy-3-methyldodecylidene)-2,2-dimethyl-1,3-dioxane-4,6-dione **3.57** (556 mg, 1.63 mmol) was dissolved in anhydrous methanol (10 mL), and the reaction mixture was heated to reflux overnight. After cooling to room temperature, the solvent was removed *in vacuo*, and the crude residue was purified by flash chromatography eluting with 0-20% ethyl acetate in hexanes to afford methyl 5-methyl-3-oxotetradecanoate **3.58** as a colorless oil composed of a mixture of keto and enol tautomers (375 mg, 1.39 mmol, 85%). TLC: R_f = 0.92 (silica gel, 1:1 hexanes:ethyl acetate). HRMS (ESI): calc'd for $C_{16}H_{31}O_3^+$ [M+H]$^+$, 271.2268; found 271.2257. ^{1}H-NMR

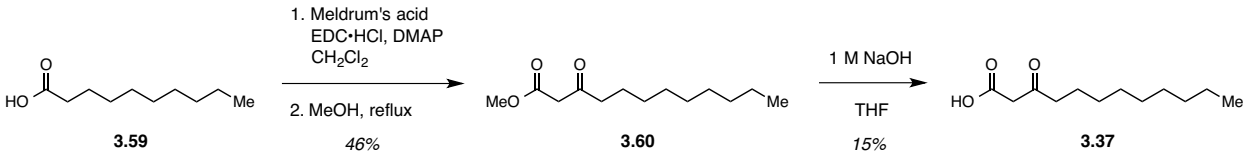
(500 MHz; $CDCl_3$): **keto form:** δ 3.76 (s, 3H, COOC**H**$_3$), 3.45 (s, 2H, CH_3OCOC**H**$_2$), 2.52 (dd, 1H, *J* = 16.5, 5.6 Hz, COC**H**$_2$), 2.34 (dd, 1H, *J* = 16.5, 8.0 Hz, COC**H**$_2$), 2.05–1.95 (m, 1H, C**H**), 1.21–1.35 (m, 16H, C**H**$_2$), 0.96–0.86 (m, 6H, C**H**$_3$); **enol form:** δ 12.02 (s, 1H, CHCO**H**), 4.98 (s, 1H, C**H**COH), 2.41 (dd, 1H, *J* = 15.8, 5.7 Hz, COC**H**$_2$), 2.25–2.19 (dd, 1H, *J* = 15.8, 6.3 Hz, COC**H**$_2$), 2.05–1.95 (m, 1H, C**H**), 1.21–1.35 (m, 16H, C**H**$_2$), 0.96–0.86 (m, 6H, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 202.5, 178.3, 173.0, 167.6, 89.8, 52.2, 51.3, 50.4, 49.5, 42.8, 37.0, 36.8, 32.0, 30.9, 29.8 (2), 29.7 (2), 29.4, 29.0, 27.0, 22.7, 19.8, 19.7, 14.1.



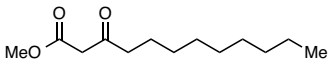
3.34

5-Methyl-3-oxotetradecanoic acid (3.34): Methyl 5-methyl-3-oxotetradecanoate **3.58** (50 mg, 0.185 mmol) was dissolved in isopropanol (2 mL), and 1 M aqueous sodium hydroxide solution (2 mL) was added to the reaction flask. The reaction mixture was stirred at room temperature for 3 h. The reaction mixture was quenched with the addition of 1 M aqueous hydrochloric acid (5 mL), and the aqueous layer was extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The residue was purified by flash chromatography, eluting with 40–60% ethyl acetate in hexanes to afford 5-methyl-3-oxotetradecanoic acid **3.34** as a colorless oil composed of a mixture of keto and enol tautomers (27 mg, 0.105 mmol, 57%). TLC: R_f = 0.12 (silica gel, 1:1 hexanes:ethyl acetate). HRMS (ESI): calc'd for $C_{15}H_{27}O_3^-$ $[M-H]^-$, 255.1966; found, 255.1963. ^{1}H-NMR (500 MHz; $CDCl_3$): **keto form:** δ 3.50 (s, 2H, CH_3OCOC**H**$_2$), 2.55 (dd, 1H, *J* = 16.5, 5.6 Hz, COC**H**$_2$), 2.36 (dd, 1H, *J* = 16.5, 8.0 Hz, COC**H**$_2$), 2.05–2.01 (m, 1H, C**H**), 1.31-1.26 (m, 19H, C**H**$_2$), 0.92 (d, 3H, *J* = 6.6 Hz, CHC**H**$_3$), 0.88 (t, 3H, *J* = 6.9 Hz, C**H**$_3$); **enol form:** δ 11.87 (s, 1H, CHCO**H**), 5.02 (s, 1H, C**H**COH), 2.24 (dd, 1H, *J* = 13.6, 5.9 Hz, COC**H**$_2$), 2.05–2.01 (m, 1H, C**H**), 1.98 (dd, 1H, *J* = 13.7, 8.3 Hz, COC**H**$_2$), 1.31-1.26 (m, 19H, C**H**$_2$), 0.92 (d, 3H, *J* = 6.6 Hz, CHC**H**$_3$), 0.88 (t, 3H, *J* = 6.9 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 204.5, 181.5, 176.6, 171.2, 89.4, 70.4, 61.7, 50.8, 48.4, 43.1, 36.9, 32.0, 31.2, 29.9, 29.8, 29.5, 29.2, 27.1, 22.8, 19.9, 19.6, 14.3.

3.4.16. Synthesis of the β-keto acid internal standard for the quantitative LC-MS assay



Scheme 3.16: Synthesis of the β-keto acid **3.37** as an internal standard for the quantitative analysis of the CylH and CylI LC-MS assays.

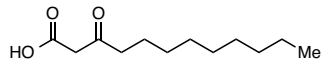


3.60

Methyl 3-oxododecanoate (3.60): To a solution of decanoic acid **3.59** (200 mg, 1.16 mmol) in dichloromethane (9 mL) at 0 ºC was added 2,2-dimethyl-1,3-dioxane-4,6-dione (167 mg, 1.16 mmol) and 4-dimethylaminopyridine (426 mg, 3.48 mmol), followed by the dropwise addition of a solution of EDC•HCl (245 mg, 1.28 mmol) in dichloromethane (3 mL) over 5 min. The reaction mixture was slowly warmed up to room temperature and stirred overnight. The reaction mixture was diluted with dichloromethane (30 mL), and the organic layer was washed with 1 M aqueous hydrochloric acid (3 x 30 mL) and brine (30 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to afford 5-(1-hydroxydecylidene)-2,2-dimethyl-1,3-dioxane-4,6-dione as a yellow oil. The crude product was used directly in the next reaction without further purification. TLC: R_f = 0.35 (silica gel, 1:1 hexanes:ethyl acetate). HRMS (ESI): calc'd for $C_{16}H_{25}O_5^-$ $[M-H]^-$, 297.1707; found, 297.1704.

Crude 5-(1-hydroxydecylidene)-2,2-dimethyl-1,3-dioxane-4,6-dione (1.16 mmol) was dissolved in anhydrous methanol (10 mL), and the reaction mixture was heated to reflux overnight. The reaction mixture was cooled to room temperature, concentrated *in vacuo* and purified by flash chromatography, eluting with 0–20% ethyl acetate in hexanes to afford methyl 3-oxododecanoate **3.60** as a colorless oil composed of a mixture of keto and enol tautomers (120 mg, 0.53 mmol, 46%). TLC: R_f = 0.71 (silica gel, 1:1 hexanes:ethyl acetate). HRMS (ESI): calc'd for $C_{13}H_{25}O_3^+$ $[M+H]^+$, 229.1798; found, 229.1817.

^{1}H and ^{13}C NMR data matched previously reported values.[53]

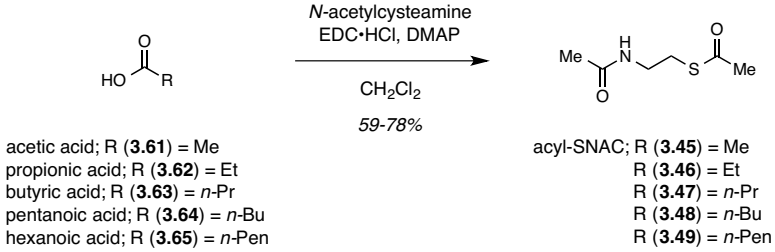


3.37

3-Oxododecanoic acid (3.37): Methyl 3-oxotetradecanoate **3.60** (35 mg, 0.15 mmol) was dissolved in THF (2 mL), and 1 M aqueous doium hydroxide solution (2 mL) was added to the reaction flask. The reaction mixture was stirred at room temperature for 3 h, and then the THF was removed *in vacuo*. The resulting mixture was diluted with ethyl acetate (5 mL) and acidified with 1 M aqueous hydrochloric acid (5 mL). The layers were separated and the aqueous layer was extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude residue was purified by flash chromatography, eluting with 50–100% ethyl acetate in hexanes to give 3-oxododecanoic acid **3.37** as mixture of keto and enol tautomers (5.0 mg, 0.023 mmol, 15%). TLC: R_f = 0.28 (silica gel, 1:1 hexanes:ethyl acetate). HRMS (ESI): calc'd for $C_{12}H_{21}O_3^-$ $[M-H]^-$, 213.1496; found 213.1498.

^{1}H and ^{13}C NMR data matched previously reported values.[53]

3.4.17. Synthesis of acyl-SNAC substrates for the spectrophotometric hydrolysis assay



Scheme 3.17: Synthesis of the short-chain acyl-SNAC substrates for the spectrophotometric hydrolysis assays.

General procedure for the synthesis of acyl-SNAC substrates (3.45-3.49): To the free acid **3.61**-**3.65** (1.03 mmol) in a flask cooled to 0 ºC, a solution of EDC•HCl (360 mg, 1.88 mmol) in dichloromethane (3 ml) was added, and the reaction mixture was stirred at 0 ºC for 20 min. To the reaction flask, *N*-acetylcysteamine (100 µL, 0.94 mmol) and 4-dimethylaminopyridine (11.5 mg, 0.094 mmol) were added.

The reaction mixture was warmed to room temperature and stirred for an additional 3 hours. The reaction mixture was quenched with the addition of water (10 mL) and extracted with ethyl acetate (3 x 15 mL). The combined organic layers were washed with 1 M aqueous hydrochloric acid (30 mL), saturated aqueous solution of sodium bicarbonate (30 mL) and brine (30 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to afford desired acyl-SNAC **3.45**-**3.49**. The isolated yields were as follows: acetyl-SNAC **3.45** (100 mg, 0.62 mmol, 60%), propionyl-SNAC **3.46** (97 mg, 0.55 mmol, 59%), butyryl-SNAC **3.47** (135 mg, 0.71 mmol, 76%), pentanoyl-SNAC **3.48** (138 mg, 0.68 mmol, 72%), and hexanoyl-SNAC **3.49** (159 mg, 0.73 mmol, 78%).

^{1}H and ^{13}C NMR data matched previously reported values.[54]

3.5. References

(1) Nakamura, H.; Wang, J. X.; Balskus, E. P. Assembly line termination in cylindrocyclophane biosynthesis: Discovery of an editing type II thioesterase domain in a type I polyketide synthase. *Chem. Sci.* **2015**, *6*, 3816.

(2) Staunton, J.; Weissman, K. J. Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.* **2001**, *18*, 380.

(3) Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part A—erythromycin biosynthesis). *Nat. Prod. Rep.* **2001**, *18*, 190.

(4) Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part B). *Nat. Prod. Rep.* **2001**, *18*, 231.

(5) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468.

(6) Du, L. C.; Lou, L. L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **2010**, *27*, 255.

(7) Bailey, A. M.; Cox, R. J.; Harley, K.; Lazarus, C. M.; Simpson, T. J.; Skellam, E. Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: First observation of a reductive release mechanism during polyketide biosynthesis. *Chem. Commun.* **2007**, 4053.

(8) Zabala, A. O.; Xu, W.; Chooi, Y. H.; Tang, Y. Characterization of a silent azaphilone gene cluster from *Aspergillus niger* ATCC 1015 reveals a hydroxylation-mediated pyran ring formation. *Chem. Biol.* **2012**, *19*, 1049.

(9) Fujii, I.; Watanabe, A.; Sankawa, U.; Ebizuka, Y. Identification of claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. *Chem. Biol.* **2001**, *8*, 189.

(10) Minto, R. E.; Townsend, C. A. Enzymology and molecular biology of aflatoxin biosynthesis. *Chem. Rev.* **1997**, *97*, 2537.

(11) Liu, T. G.; You, D. L.; Valenzano, C.; Sun, Y. H.; Li, J. L.; Yu, Q.; Zhou, X. F.; Cane, D. E.; Deng, Z. X. Identification of nane as the thioesterase for polyether chain release in nanchangmycin biosynthesis. *Chem. Biol.* **2006**, *13*, 945.

(12) Harvey, B. M.; Hong, H.; Jones, M. A.; Hughes-Thomas, Z. A.; Goss, R. M.; Heathcote, M. L.; Bolanos-Garcia, V. M.; Kroutil, W.; Staunton, J.; Leadlay, P. F.; Spencer, J. B. Evidence that a novel thioesterase is responsible for polyketide chain release during biosynthesis of the polyether ionophore monensin. *ChemBioChem* **2006**, *7*, 1435.

(13) Liu, T. G.; Lin, X.; Zhou, X. F.; Deng, Z. X.; Cane, D. E. Mechanism of thioesterase-catalyzed chain release in the biosynthesis of the polyether antibiotic nanchangmycin. *Chem. Biol.* **2008**, *15*, 449.

(14) Xie, X. K.; Meehan, M. J.; Xu, W.; Dorrestein, P. C.; Tang, Y. Acyltransferase mediated polyketide release from a fungal megasynthase. *J. Am. Chem. Soc.* **2009**, *131*, 8388.

(15) Sucipto, H.; Wenzel, S. C.; Muller, R. Exploring chemical diversity of pyrone antibiotics: Molecular basis of myxopyronin biosynthesis. *ChemBioChem* **2013**, *14*, 1581.

(16) Gerber, R.; Lou, L. L.; Du, L. C. A PLP-dependent polyketide chain releasing mechanism in the biosynthesis of mycotoxin fumonisins in *Fusarium verticillioides*. *J. Am. Chem. Soc.* **2009**, *131*, 3148.

(17) Awakawa, T.; Yokota, K.; Funa, N.; Doi, F.; Mori, N.; Watanabe, H.; Horinouchi, S. Physically discrete beta-lactamase type thioesterase catalyzes product release in atrochrysone synthesis by iterative type I polyketide synthase. *Chem. Biol.* **2009**, *16*, 613.

(18) Frank, B.; Wenzel, S. C.; Bode, H. B.; Scharfe, M.; Blocker, H.; Müller, R. From genetic diversity to metabolic unity: Studies on the biosynthesis of aurafurones and aurafuron-like structures in myxobacteria and streptomycetes. *J. Mol. Biol.* **2007**, *374*, 24.

(19) Cantu, D. C.; Chen, Y. F.; Reilly, P. J. Thioesterases: A new perspective based on their primary and tertiary structures. *Protein Sci.* **2010**, *19*, 1281.

(20) Heathcote, M. L.; Staunton, J.; Leadlay, P. F. Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **2001**, *8*, 207.

(21) Butler, A. R.; Bate, N.; Cundliffe, E. Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae. Chem. Biol.* **1999**, *6*, 287.

(22) Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14083.

(23) Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A. Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. *J. Biol. Chem.* **2002**, *277*, 48028.

(24) Yeh, E.; Kohli, R. M.; Bruner, S. D.; Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBioChem* **2004**, *5*, 1290.

(25) Kotowska, M.; Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7735.

(26) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P. Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J. Am. Chem. Soc.* **2012**, *134*, 18518.

(27) Miyanaga, A.; Funa, N.; Awakawa, T.; Horinouchi, S. Direct transfer of starter substrates from type I fatty acid synthase to type III polyketide synthases in phenolic lipid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 871.

(28) Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in *Bacillus subtilis. J. Bacteriol.* **2009**, *191*, 4916.

(29) Austin, M. B.; Saito, T.; Bowman, M. E.; Haydock, S.; Kato, A.; Moore, B. S.; Kay, R. R.; Noel, J. P. Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid - type III polyketide synthase. *Nat. Chem. Biol.* **2006**, *2*, 494.

(30) Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998**, *37*, 1585.

(31) La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. Manipulation of carrier proteins in antibiotic biosynthesis. *Chem. Biol.* **2004**, *11*, 195.

(32) Buntin, K.; Weissman, K. J.; Muller, R. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis. *ChemBioChem* **2010**, *11*, 1137.

(33) Claxton, H. B.; Akey, D. L.; Silver, M. K.; Admiraal, S. J.; Smith, J. L. Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *J. Biol. Chem.* **2009**, *284*, 5021.

(34) Whicher, J. R.; Florova, G.; Sydor, P. K.; Singh, R.; Alhamadsheh, M.; Challis, G. L.; Reynolds, K. A.; Smith, J. L. Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in *Streptomyces coelicolor*. *J. Biol. Chem.* **2011**, *286*, 22558.

(35) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: Architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421.

(36) Julien, B.; Tian, Z. Q.; Reid, R.; Reeves, C. D. Analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis. *Chem. Biol.* **2006**, *13*, 1277.

(37) Ahrendt, T.; Miltenberger, M.; Haneburger, I.; Kirchner, F.; Kronenwerth, M.; Brachmann, A. O.; Hilbi, H.; Bode, H. B. Biosynthesis of the natural fluorophore legioliulin from legionella. *ChemBioChem* **2013**, *14*, 1415.

(38) Muliandi, A.; Katsuyama, Y.; Sone, K.; Izumikawa, M.; Moriya, T.; Hashimoto, J.; Kozone, I.; Takagi, M.; Shin-ya, K.; Ohnishi, Y. Biosynthesis of the 4-methyloxazoline-containing nonribosomal peptides, JBIR-34 and -35, in *Streptomyces* sp. Sp080513GE-23. *Chem. Biol.* **2014**, *21*, 923.

(39) Zou, Y.; Yin, H. X.; Kong, D. K.; Deng, Z. X.; Lin, S. J. A trans-acting ketoreductase in biosynthesis of a symmetric polyketide dimer SIA7248. *ChemBioChem* **2013**, *14*, 679.

(40) Yin, X.; Zabriskie, T. M. The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* **2006**, *152*, 2969.

(41) Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A. Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J. Am. Chem. Soc.* **2012**, *134*, 6865.

(42) Hou, J.; Robbel, L.; Marahiel, M. A. Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture. *Chem. Biol.* **2011**, *18*, 655.

(43) He, H. Y.; Tang, M. C.; Zhang, F.; Tang, G. L. Cis-double bond formation by thioesterase and transfer by ketosynthase in FR901464 biosynthesis. *J. Am. Chem. Soc.* **2014**, *136*, 4488.

(44) Gehret, J. J.; Gu, L. C.; Gerwick, W. H.; Wipf, P.; Sherman, D. H.; Smith, J. L. Terminal alkene formation by the thioesterase of curacin A biosynthesis structure of a decarboxylating thioesterase. *J. Biol. Chem.* **2011**, *286*, 14445.

(45) Gaudelli, N. M.; Townsend, C. A. Epimerization and substrate gating by a TE domain in beta-lactam antibiotic biosynthesis. *Nat. Chem. Biol.* **2014**, *10*, 251.

(46) Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647.

(47) Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **2006**, *1*, 280.

(48) Beltran-Alvarez, P.; Cox, R. J.; Crosby, J.; Simpson, T. J. Dissecting the component reactions catalyzed by the actinorhodin minimal polyketide synthase. *Biochemistry* **2007**, *46*, 14672.

(49) Soding, J.; Biegert, A.; Lupas, A. N. The HHPred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244.

(50) Huelsenbeck, J. P.; Ronquist, F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754.

(51) Kotaka, M.; Kong, R.; Qureshi, I.; Ho, Q. S.; Sun, H. H.; Liew, C. W.; Goh, L. P.; Cheung, P.; Mu, Y. G.; Lescar, J.; Liang, Z. X. Structure and catalytic mechanism of the thioesterase CalE7 in enediyne biosynthesis. *J. Biol. Chem.* **2009**, *284*, 15739.

(52) Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S. N.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.;

Thanki, N.; Wang, Z. X.; Yamashita, R. A.; Zhang, D. C.; Zheng, C. J.; Bryant, S. H. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222.

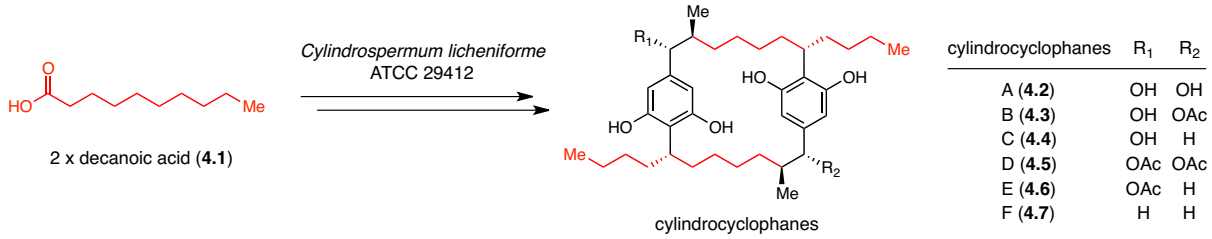
(53) Kwan, E. E.; Scheerer, J. R.; Evans, D. A. The stereochemical course of intramolecular Michael reactions. *J. Org. Chem.* **2013**, *78*, 175.

(54) Prasad, G.; Borketey, L. S.; Lin, T. Y.; Schnarr, N. A. A mechanism-based fluorescence transfer assay for examining ketosynthase selectivity. *Org. Biomol. Chem.* **2012**, *10*, 6717.

Chapter 4. Investigation of the C–H activation step in the cylindrocyclophane biosynthesis

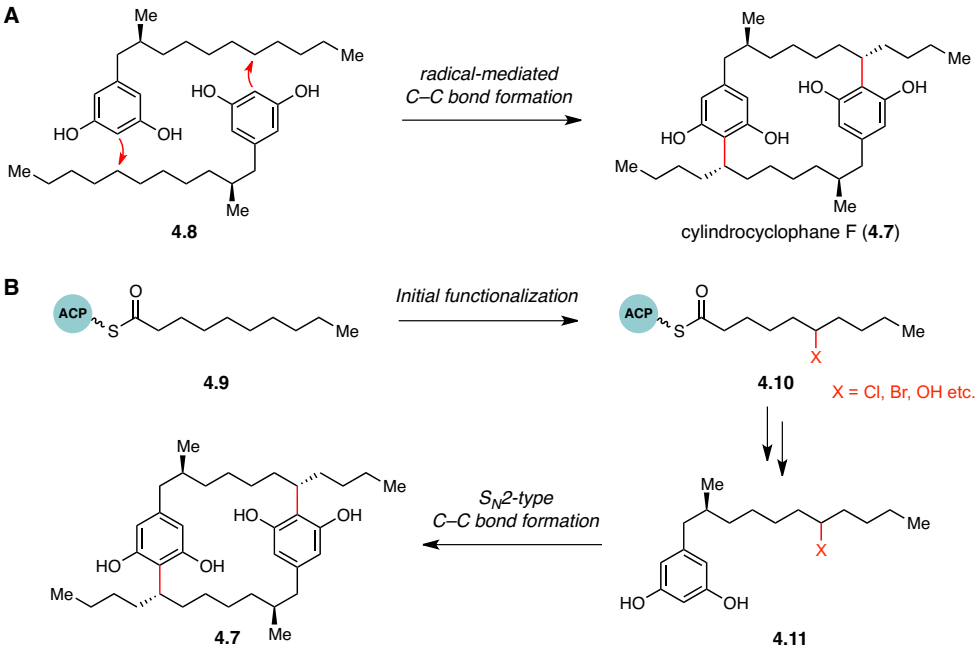
4.1. Introduction

An important step in the cylindrocyclophane biosynthesis is the functionalization of the unactivated aliphatic carbon center. As detailed in **Chapter 2**, feeding of d_{19}-decanoic acid to the native producer indicated that decanoic acid (**4.1**, **Scheme 4.1**) is the precursor to the cylindrocyclophanes (**4.2**-**4.7**) and revealed that their biosynthesis involves functionalization of an unactivated aliphatic carbon center to form the dimeric paracyclophane scaffold.[1] We considered two possible routes for how the aliphatic carbon center activation and C–C bond formation may proceed: **A**) the direct alkylation route and **B**) the pre-functionalization route (**Scheme 4.2**).



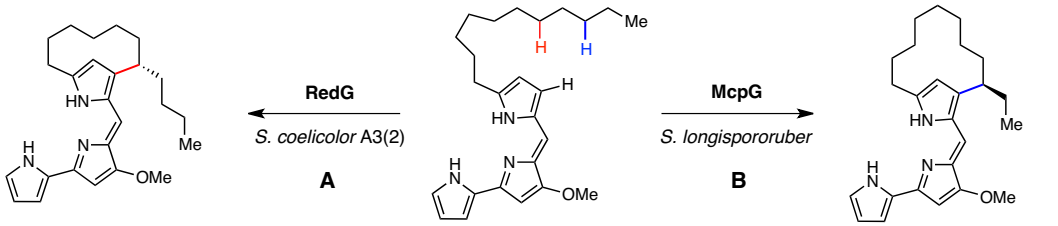
cylindrocyclophanes	R_1	R_2
A (**4.2**)	OH	OH
B (**4.3**)	OH	OAc
C (**4.4**)	OH	H
D (**4.5**)	OAc	OAc
E (**4.6**)	OAc	H
F (**4.7**)	H	H

Scheme 4.1: Incorporation of decanoic acids into cylindrocyclophane scaffold shows that the cylindrocyclophane biosynthesis requires functionalization of the aliphatic carbon center to construct the [7.7]paracyclophane scaffold.



Scheme 4.2: Two possible routes for the dimerization of the monomeric resorcinol units. **A**) Direct alkylation route. **B**) Pre-functionalization route.

The direct alkylation route (**Scheme 4.2A**) involves a radical-mediated process in which one enzyme is responsible for both the activation of the aliphatic carbon center and the C–C bond formation. Direct oxidative alkylations are rare in nature, but several enzymes have been observed to catalyze such transformations.[2-4] The first example of an enzyme-catalyzed direct oxidative alkylation occurs in the biosynthesis of the prodiginines and involves the Rieske oxygenases RedG and McpG (**Scheme 4.3**).[2] Rieske oxygenases are often associated with oxidative degradation of aromatic compounds, and these enzymes contain a [2Fe–2S] cluster as well as a catalytic non-heme iron center that coordinates the iron cofactor with a "2-His-1-carboxylate triad" (two histidine residues and an aspartate/glutamate residue).[5] Unlike other characterized Rieske oxygenases that catalyze mono- or dioxygenations, RedG and McpG catalyze regio- and stereoselective oxidative cyclization of undecylprodigiocin (**4.12**) to form 10- and 12-membered carbocyclic rings in the antibiotics streptorubin B (**4.13**) and metacycloprodigiosin (**4.14**), respectively.[2]



Scheme 4.3: Oxidative C–C bond formation catalyzed by Rieske oxygenases **A**) RedG and **B**) McpG.

The function of RedG was confirmed by gene deletion and complementation experiments in the native producer, *Streptomyces coelicolor* A3(2), as well as feeding studies in a heterologous expression host expressing RedG. In addition, replacement of *redG* with *mcpG* in *S. coelicolor* A3(2) resulted in the production of metacycloprodigiosin (**4.14**) instead of streptorubin B (**4.13**).[2] Interestingly, the observed stereochemistry of the RedG (**Scheme 4.3A**) and McpG (**Scheme 4.3B**) catalyzed C–C bond formations are different, indicating that these homologous enzymes have opposite stereoselectivity and the cyclization step leads to both regio- and stereodivergence. More recently, the C–H activation and C–C bond formation catalyzed by RedG and McpG have been shown to proceed through inversion of stereochemistry by additional feeding studies using synthetic substrate analogs.[6] Homologs of RedG and McpG are also thought to catalyze similar sp^2–sp^3 C–C bond forming reactions in the biosyntheses of other prodiginines such as prodigiosin R1, roseophilin and marineosins.[2,7] While the functions of these Rieske oxygenases have been confirmed *in vivo*, biochemical characterizations of these enzymes are yet to be reported.

In addition to the Rieske oxygenases, two other enzymes, symerythrin[3] and StrB,[4] have recently been found to catalyze direct oxidative alkylations to generate sp^2–sp^3 C–C bonds. Symerythrin is an enzyme that belongs to the ferritin-like superfamily, and this enzyme is known to form a phenylalanine–valine crosslink in its active site (**4.15**, see **Scheme 4.4A**).[3] While the physiological function of symerythrin is unknown, the phenylalanine–valine crosslink generated by the di-iron cofactor in its active site has been shown to stabilize the protein structure.[3] StrB belongs to a family of radical-SAM enzymes, and radical-

SAM enzymes are known to catalyze diverse set of radical reactions, including C–C bond formation.[4,8] The radical SAM enzyme StrB generates a lysine–tryptophan crosslink (**Scheme 4.4B**) in the biosynthesis of streptide (**4.16**).[4] Both symerythrin and StrB catalyze C–C bond formation between an unactivated carbon center and an aromatic carbon center. A similar C–C bond formation could take place in cylindrocyclophane biosynthesis if the pathway proceeds through a directly oxidative alkylation route.

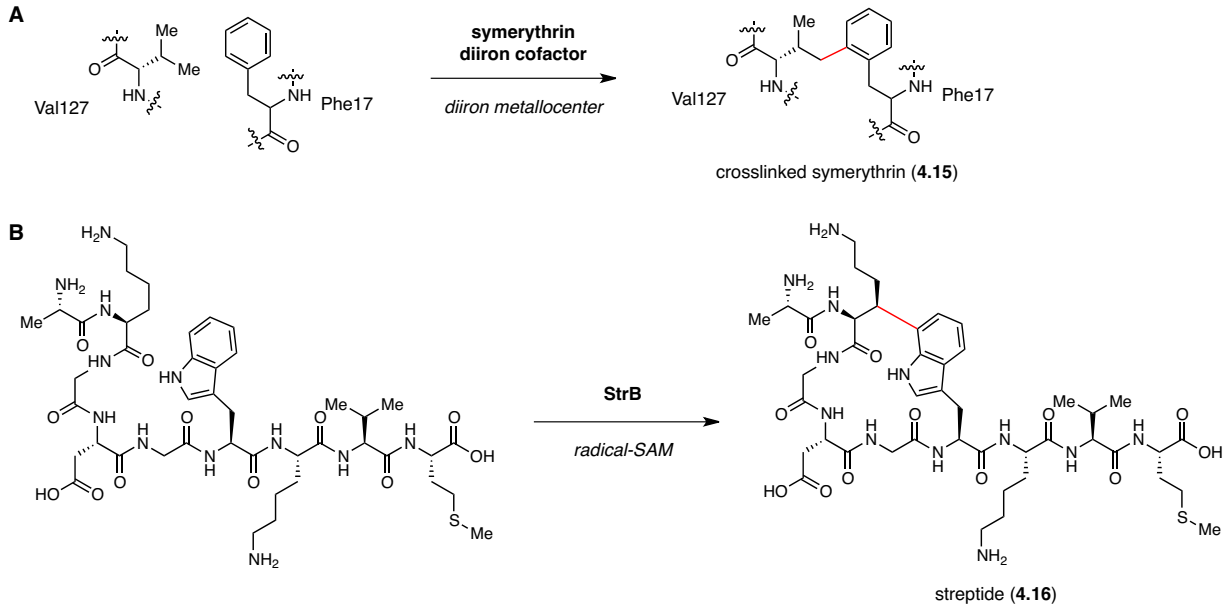
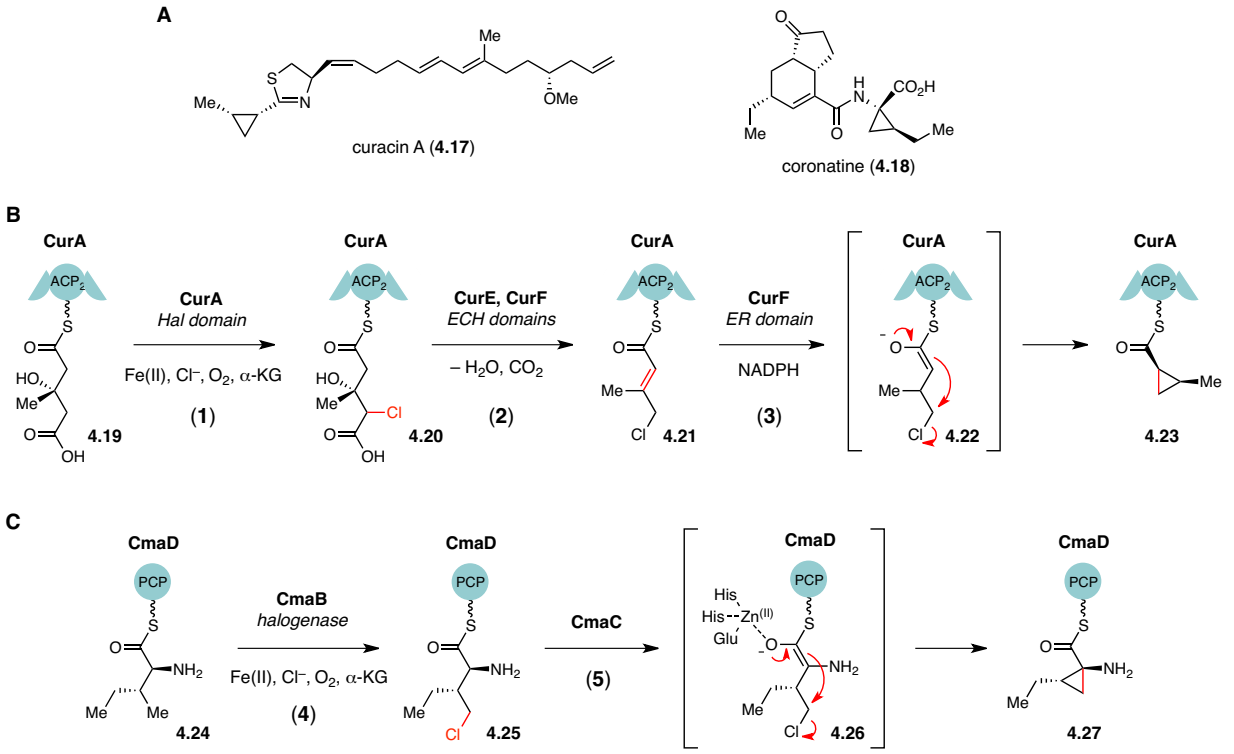


Figure 4.4: Other examples of enzymatic direct alkylations. **A**) Val–Phe crosslink generated by the di-iron enzyme symerythrin. **B**) Lys–Trp crosslink generated by the radical-SAM enzyme StrB.

While the direct alkylation route would involve only one enzyme, the pre-functionalization route (**Scheme 4.2B**) would likely require one enzyme for the initial functionalization of the aliphatic carbon center and another enzyme for the C–C bond formation between two pre-functionalized monomeric resorcinol units (**4.11**). This type of C–C bond formation logic is found in the biosyntheses of natural products containing a cyclopropane moiety. For instance, cyclopropanation has been characterized for the biosyntheses of curacin (**4.17**, see **Scheme 4.5A**)[9] and coronatine (**4.18**).[10] Both pathways employ analogous logic to generate the cyclopropane moiety through cryptic chlorination followed by an intramolecular S_N2 reaction that involves an enolate nucleophile and an alkyl chloride electrophile. In both cases, the chlorination step (**Reaction 1** in **Scheme 4.5B** and **Reaction 4** in **Scheme 4.5C**) is

catalyzed by a α-ketoglutarate-dependent non-heme halogenase that exists either as a domain on an assembly line enzyme or as a discrete enzyme. In the curacin (**4.17**) biosynthesis, the ER domain catalyzes the subsequent C–C bond formation through reduction of the α,β-unsaturated thioester (**4.21**) using NADPH (**Reaction 3**, **Scheme 4.5B**), and the resulting enolate (**4.22**) displaces the chlorine to form the cyclopropane ring (**4.23**).[11] On the other hand, the C–C bond formation in the coronatine (**4.18**) biosynthesis is catalyzed by the Zn-dependent enzyme CmaC with homology to methylmalonyl-CoA epimerases (**Reaction 5**, **Scheme 4.5C**).[9,11] While we can imagine similar pre-functionalization and C–C bond formation steps to take place for the cylindrocyclophane biosynthesis, notable differences are that the nucleophilic attack must occur on a secondary carbon instead of a primary carbon and that the alkylation is intermolecular and involves an aromatic nucleophile.



Scheme 4.5: The cyclopropane moiety in curacin and coronatine is constructed through cryptic chlorination followed by S_N2 reaction. **A**) Structures of curacin A and coronatine. **B**) Cyclopropanation in the curacin biosynthesis. **C**) Cyclopropanation in the coronatine biosynthesis.

To the best of our knowledge, all the known examples of the C–C bond formation for both the direct alkylation and the pre-functionalization routes are intramolecular reactions.[2-4,10,11] In the case of the cylindrocyclophanes, the two C–C bond formation is required to construct the paracyclophane scaffold. These reactions might be sequential, which requires the first C–C bond formation to be intermolecular and the second C–C bond formation to be intramolecular (**Scheme 4.2**). In addition, the conserved [7.7]paracyclophane structures of the cylindrocyclophanes (**4.2**-**4.7**)[12] and their analogs[13-17] imply that the key C–C bond formation in the biosynthesis of these molecules is both regio- and stereoselective. Understanding how enzymes can catalyze such a challenging transformation with high selectivity is stimulating from both chemical and biological perspectives.

Both the direct alkylation and the pre-functionalization routes require a metalloenzyme for the initial activation of the aliphatic C–H bond. With the assumption that all the enzymes required for the cylindrocyclophane biosynthesis are present in the *cyl* gene cluster (**Figure 4.1**), we searched for the enzymes that may be capable of functionalizing the aliphatic carbon center. The two candidate enzymes we considered were CylC and CylP. Sequence and secondary structural homology search using protein BLAST and HHPred revealed that CylP is likely a Rieske oxygenase that has a weak homology to RedG[2] (15% identity and 29% similarity). Due to the precedence of the Rieske oxygenases involvement in oxidative C–C bond formations,[2] we thought that CylP was the best candidate enzyme for the direct alkylation route. The other candidate enzyme, CylC, does not have a primary sequence homology to previously characterized proteins and was consequently annotated as a "hypothetical protein". Nevertheless, further bioinformatic analyses indicate that CylC is likely a di-metallo carboxylate enzyme, which might be responsible for catalyzing the direct alkylation or the functionalization step in cylindrocyclophane biosynthesis. Further bioinformatic analyses indicate that CylC is a new class of halogenase that catalyzes a cryptic chlorination reaction. CylC, therefore, is the candidate enzyme that catalyzes the initial functionalization step in the pre-functionalization route.

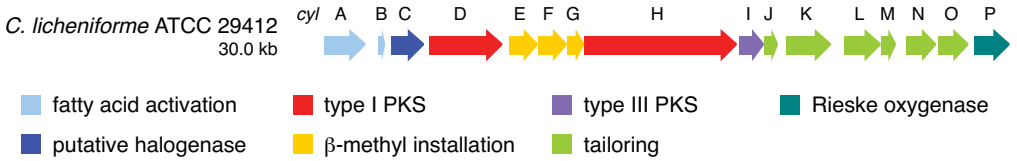


Figure 4.1: The cylindrocyclophane (*cyl*) biosynthetic gene cluster.

4.2. Results and Discussions

4.2.1. Investigation of the CylP activity

To determine the function of CylP, we initially attempted to overexpress and purify the enzyme through heterologous expression in *E. coli.* Our attempts to purify the enzyme, however, were not fruitful. The difficulty in obtaining CylP as a soluble protein may be due to the presence of two predicted transmembrane regions at the C-terminus of the protein based on several transmembrane prediction tools including DAS,[18] PRED-TMR,[19] TMHMM[20] and Phobius.[21] Consequently, we relied on feeding experiments in an *E. coli* strain that overexpresses CylP with an N-terminal maltose binding protein tag (MBP-CylP) for the initial characterization of this enzyme.

In these experiments, the racemic synthetic resorcinol (**4.28**, synthesis described in **Chapter 2**) was added to the cultures of *E. coli* expressing MBP-CylP (**Figure 4.2A**). A strain of *E. coli* transformed with an empty vector that expresses just the MBP tag was used as a negative control. After incubation of the *E. coli* cultures with resorcinol substrate (**4.28**), the whole cultures were extracted, and the organic extracts were analyzed by LC-HRMS. A mass corresponding to cylindrocyclophane F (**4.7**), which results from the dimerization of the resorcinol substrates, was not observed in these culture extracts. Instead, a new peak corresponding to the mass of hydroxylated resorcinol (**4.29**) was observed only in the cultures of *E. coli* containing MBP-CylP. MS/MS analysis of this peak showed that the hydroxylation occurred at the benzylic position of the resorcinol (**Figure 4.2C**). Thus, the feeding experiment revealed that CylP is catalyzing benzylic hydroxylation, instead of the expected C–C bond formation, to install the functional group required for the production of cylindrocyclophanes A–E.

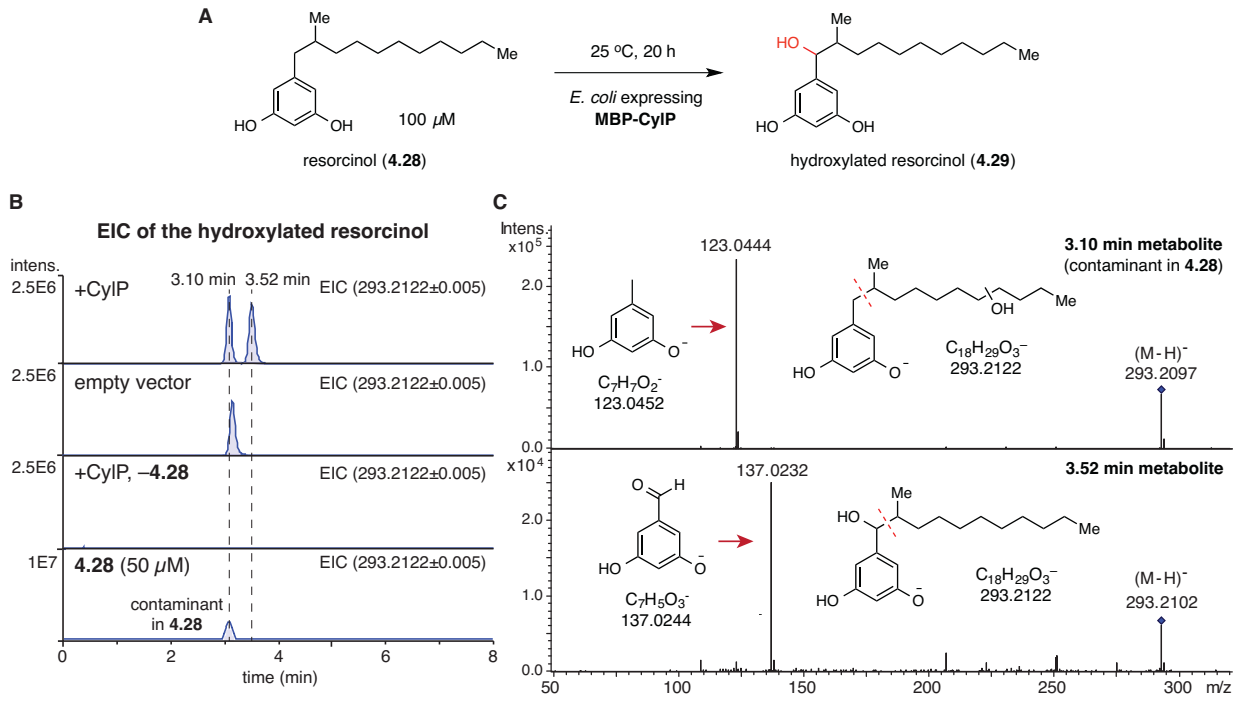
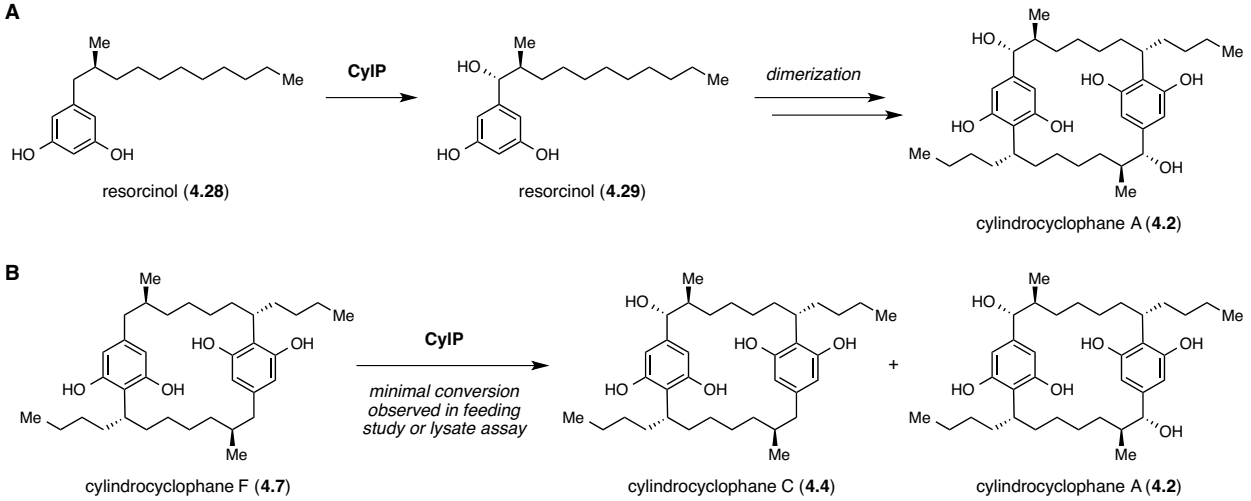


Figure 4.2: CylP catalyzes benzylic hydroxylation of resorcinol **4.28**. **A**) Feeding experiment using *E. coli* strain expressing MBP-CylP. **B**) LC-MS extracted ion chromatograms for the mass of hydroxylated resorcinol. **C**) MS/MS analysis shows that the CylP-catalyzed hydroxylation of resorcinol **4.28** occurs at the benzylic position.

While we determined that CylP installs the benzylic hydroxyl group for the production of cylindrocyclophanes A-E (**1**-**5**), the timing of this tailoring event was uncertain. CylP may not only use the monomeric resorcinol (**4.28**) but also the dimeric cylindrocyclophane F (**4.7**) as a substrate (**Scheme 4.6**). To investigate the timing of the CylP-catalyzed hydroxylation, we repeated the same feeding experiment as above using cylindrocyclophane F (**4.7**) as the substrate. LC-HRMS analysis showed that the masses corresponding to cylindrocyclophanes A (**4.2**) and C (**4.4**) were absent in the culture extracts. The result could indicate that CylP does not use cylindrocyclophane F as a substrate and that the hydroxylation step occurs prior to the dimerization step (**Scheme 4.6A**). Alternatively, the lack of CylP activity on cylindrocyclophane F in the feeding study may be due to the membrane impermeability of the substrate. To rule out this possibility, we developed an assay using the lysate of the *E. coli* culture expressing MBP-CylP. In the lysate assays, we were able to see the conversion of the resorcinol (**4.28**) to the hydroxylated resorcinol (**4.29**, see **Figure 4.3A**), but minimal conversion of cylindrocyclophane F

(**4.7**) to cylindrocyclophanes A (**4.2**) and C (**4.4**, see **Figure 4.3B**). The lysate assay result confirmed that CylP prefers to use monomeric substrate, which implies that the hydroxylation takes place before the dimerization. This finding was unexpected, since the monomeric resorcinol (**4.28**) did not appear to accumulate in the extracts from the native producer, *C. licheniforme* ATCC 29412, and the feeding of synthetic resorcinol (**4.28**) resulted in cell death.[1] In conclusion, the Rieske oxygenase CylP is a hydroxylase and it does not catalyze an oxidative C–C bond formation in the dimerization step.



Scheme 4.6: Possible timing of the CylP-catalyzed benzylic hydroxylation and the dimerization steps. **A**) CylP-catalyzed hydroxylation prior to dimerization. **B**) CylP-catalyzed hydroxylation after dimerization.

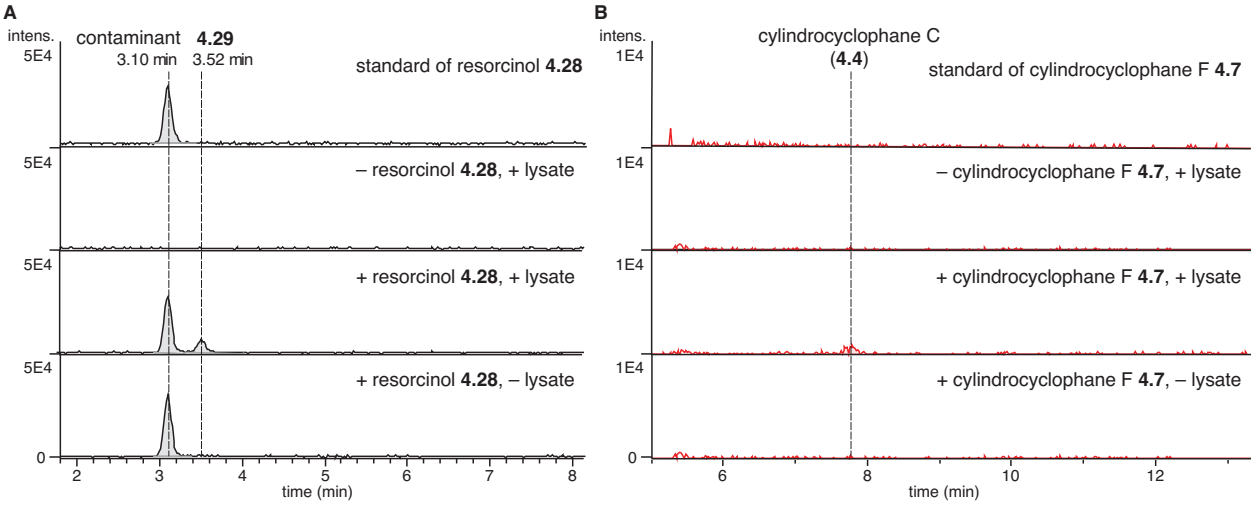
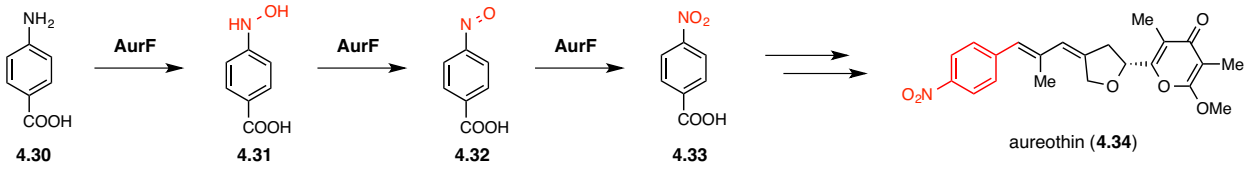


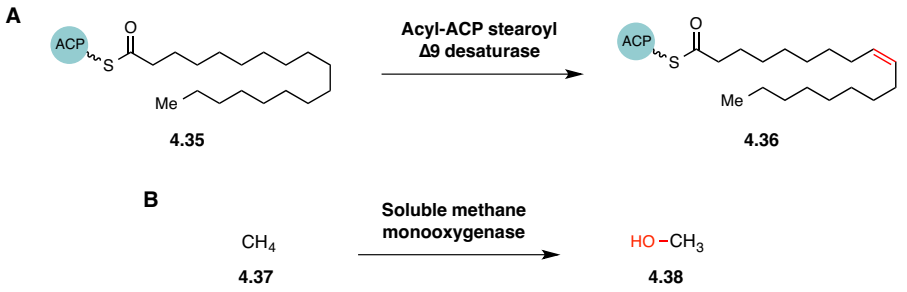
Figure 4.3: LC-MS analysis of the CylP lysate assays. **A**) Extracted ion chromatograms of the hydroxylated resorcinol mass (293.2122 ± 0.005) indicate that CylP activity can be observed using the *E. coli* lysate. **B**) Trace peak that may correspond to cylindrocyclophane C (567.4055 ± 0.005) is present in the full lysate assay, but the conversion is very poor.

4.2.2. Bioinformatic investigation of the function of CylC and its homologs

The other candidate enzyme for the functionalization of the unactivated aliphatic carbon center is CylC, which is annotated as a hypothetical protein. Because a simple primary sequence homology search did not provide information on the function of this protein, we performed a secondary structural homology search using HHPred[22] to look for possible structural similarities to known enzymes. The search result showed AurF as the best secondary structural homolog (probability = 99.8%, E-value = $1.2e^{-17}$). AurF is a di-iron enzyme that catalyzes sequential oxidation of *p*-aminobenzoic acid (**4.30**) to *p*-nitrobenzoic acid (**4.33**) in the biosynthesis of aureothin (**4.34**) in *Streptomyces thioluteus* (**Scheme 4.7**).[23] AurF belongs to the family of ferritin-like di-iron carboxylate enzymes, which have a characteristic four-helix-bundle protein structure.[24,25] Ferritin-like di-iron carboxylate enzymes bind and activate O_2 to catalyze varieties of oxidation reactions. Examples of this family of enzymes include fatty acid desaturases (**Reaction 1**, **Scheme 4.8**)[26,27] and soluble methane monooxygenases (**Reaction 2**, **Scheme 4.8**),[28] both of which are capable of functionalizing unactivated carbon centers.



Scheme 4.7: AurF-catalyzed oxidation of *p*-aminobenzoic acid to *p*-nitrobenzoic acid in the aureothin biosynthesis.



Scheme 4.8: Functionalization of unactivated carbon centers by fatty acid desaturase and methane monooxygenase.

A homology model of CylC was generated with Modeller[29] using the crystal structure of AurF containing a di-iron confactor[23] as a template. While the primary sequence homology of CylC and AurF is relatively

low (12% identity, 26% similarity), a CylC homology model can be mapped onto the four-helix-bundle structure of AurF (**Figure 4.4A**). In addition, residues that are important for binding the di-iron cofactor in AurF appear to be fully conserved in the homology model of CylC (**Figure 4.4B**). The multiple sequence alignment of CylC and its homologs showed that all seven His and Glu residues that are predicted to bind the metal cofactor are present (**Figure 4.4C**). The presence of the residues required for metal binding in CylC suggests that CylC and its homologs are likely members of a new class of ferritin-like di-metallo carboxylate enzymes.

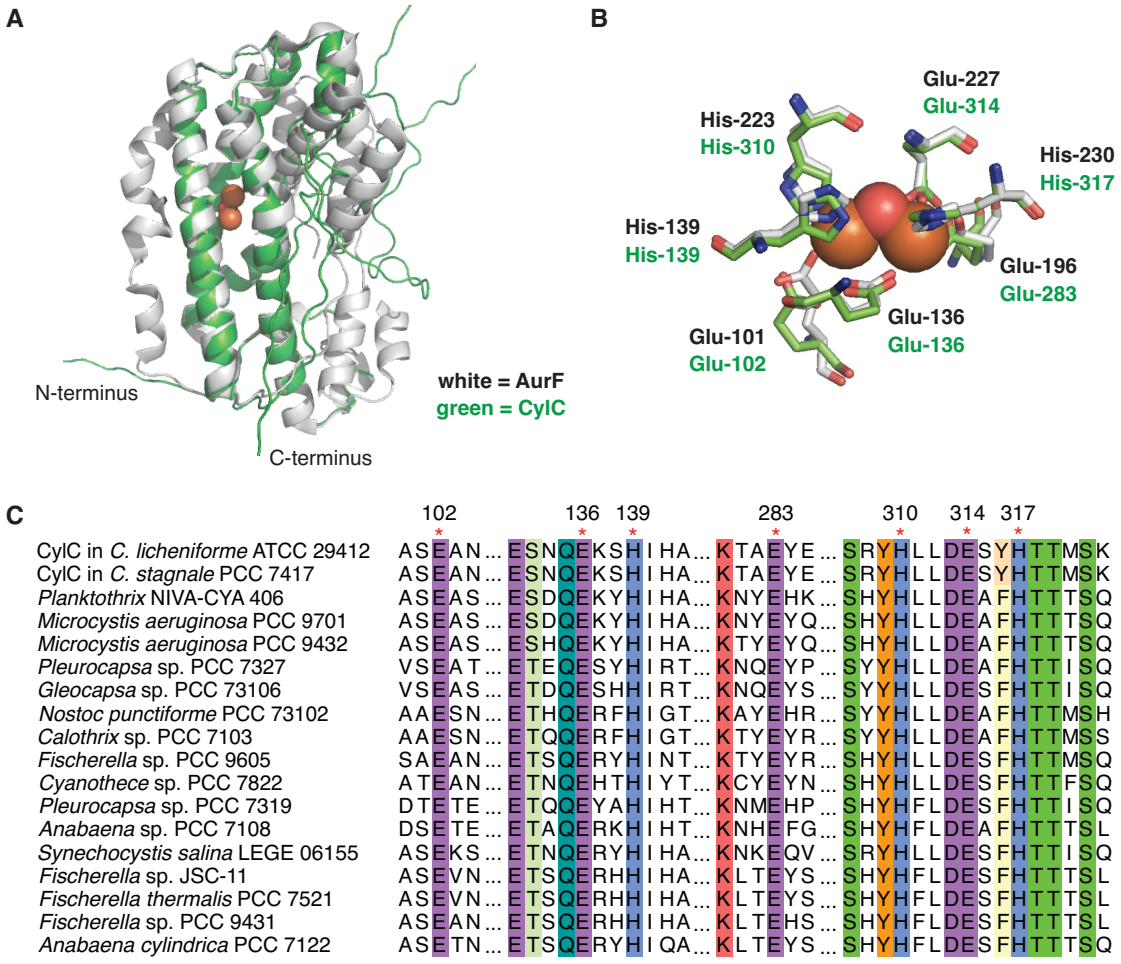


Figure 4.4: Homology model of CylC generated using AurF as the template. **A**) Overall structure of CylC mapped onto AurF. **B**) The residues required for the di-iron cofactor binding are conserved. **C**) Multiple sequence alignment of CylC and its homologs also shows conservation of the metal-binding residues (marked with *).

While this analysis suggested that CylC is likely a metalloprotein, the function of this enzyme was unknown. To obtain further insight into the possible function of CylC, we examined the genomic context of the CylC homologs in other biosynthetic gene clusters. Several cyanobacterial gene clusters with known biosynthetic products contain CylC homologs. For instance, we found a CylC homolog (50% identity, 65% similarity) that is clustered with fatty acid activating enzymes (CylA and CylB homologs) in the microginin (*mic*) biosynthetic gene cluster in *Microcystis aeruginosa* PCC 9432 (**Figure 4.5A**). Microginins (**4.39**-**4.45**, see **Figure 4.5B**) are linear *N*-acyl lipopeptides containing four to six amino acids isolated from strains of *Microcystis.*[30-32] These peptides contain a characteristic 3-amino-2-hydroxydecanoic acid (Ahda) moiety at the N-terminus, and some of the microginins such as microginin T1[33] (**4.40**) and microginins 299A (**4.42**), B (**4.43**) and D (**4.45**)[34] are known to have mono- or dichlorination of the Ahda moiety. While the biosynthetic gene cluster of microginins has been previously characterized,[35] the halogenase responsible for the chlorination of the Ahda has not been identified. The chlorination of the Ahda moiety likely requires a metalloenzyme that catalyzes radical halogenation, and we postulated that the CylC homolog in the *mic* gene cluster is responsible for catalyzing this reaction.

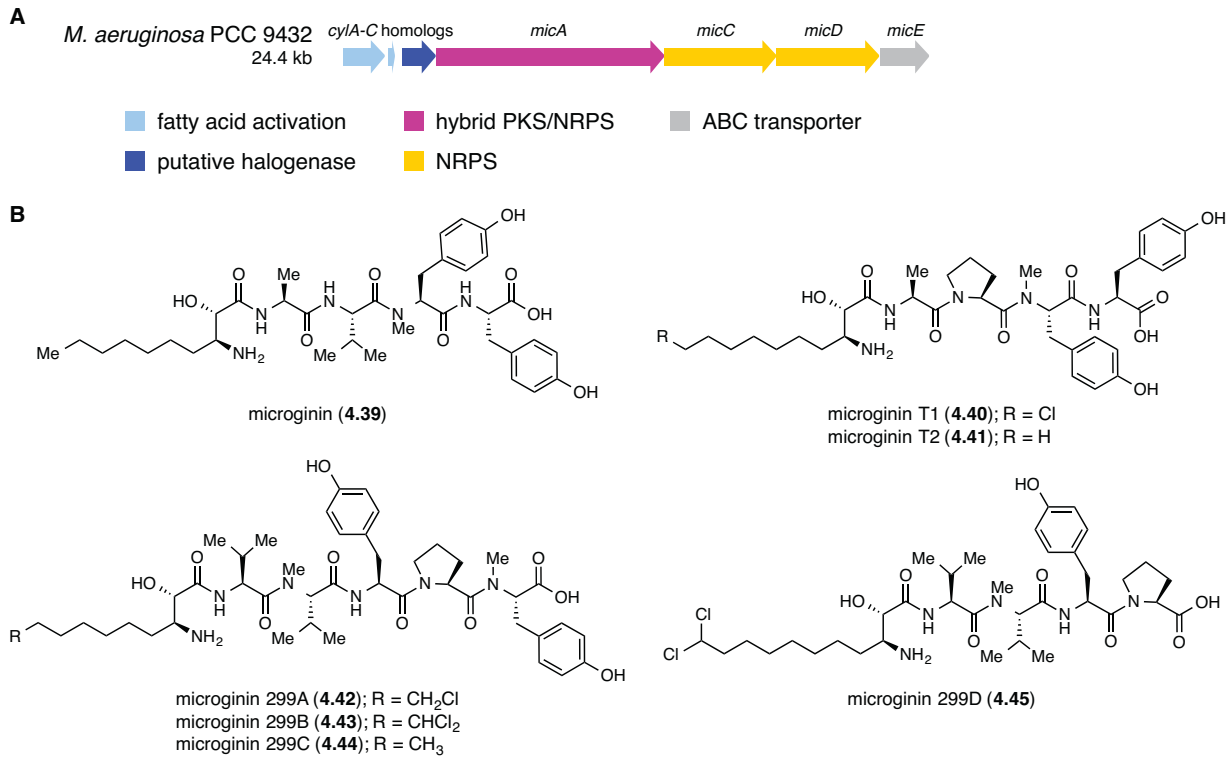


Figure 4.5: The microginin (*mic*) biosynthetic gene cluster contains a CylC homolog. **A**) The *mic* gene cluster in *Microcystis aeruginosa* PCC 9432. **B**) Structures of the microginins.

Further support for the putative halogenase function of CylC and its homologs was discovered through our effort to characterize the biosynthesis of the bartolosides (**4.46-4.49**, see **Figure 4.6A**), a family of glycosylated dialkylresorcinols (DARs) containing chlorine atoms on their lipid tails.[36] We identified the bartoloside (*brt*) biosynthetic gene cluster in *Synechocystis salina* LEGE 06155, the native producer of bartolosides B-D (**4.47-4.49**), through genome mining. Analysis of the *brt* gene cluster as well as the genome of *S. salina* LEGE 06155 revealed that the homologs of the non-heme iron and the flavin dependent halogenases are absent (**Figure 4.6B**). Instead, the *brt* gene cluster contains a CylC homolog, BrtJ (40% identity, 55% similarity), which may be responsible for the chlorination of the bartolosides.

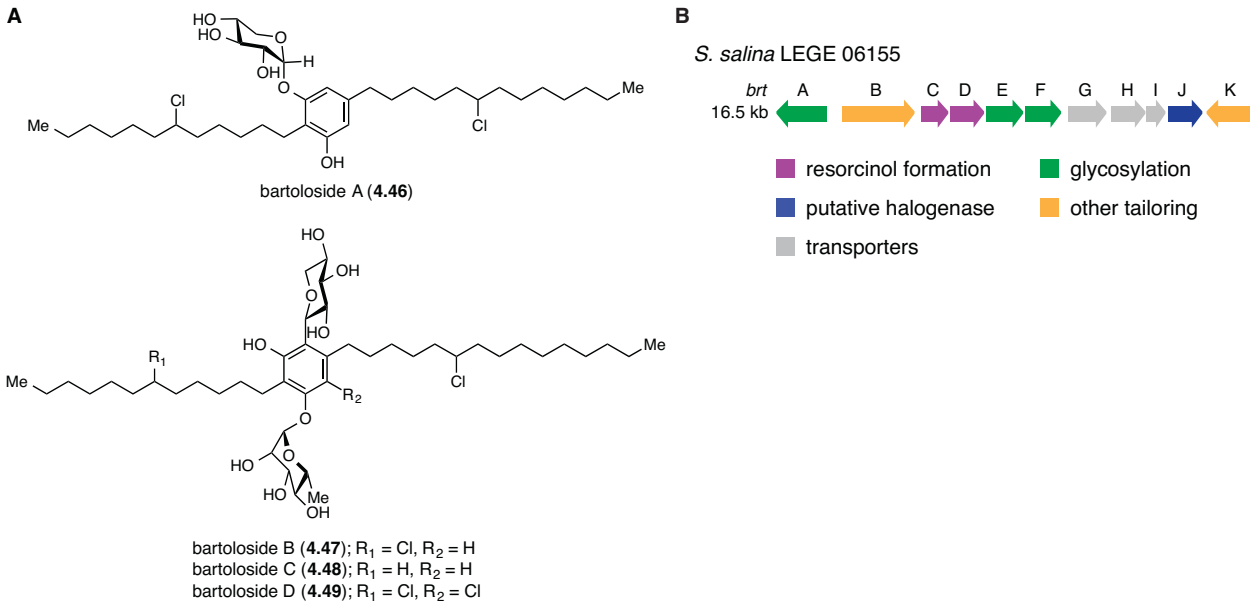


Figure 4.6: Discovery of the bartoloside (*brt*) biosynthetic gene cluster. **A**) Structures of the bartolosides. **B**) The *brt* gene cluster contains a CylC homolog, BrtJ.

More recently, the biosynthetic gene cluster of the chlorinated natural products columbamides (**4.50**-**4.52**, see **Figure 4.7A**) has been found to encode CylC homologs.[37] Interestingly, these metabolites have chlorination in the middle and at the terminus of the alkyl chain. The columbamide (*col*) biosynthetic gene cluster in *Moorea bouillonii* PNG (**Figure 4.7B**) contains two CylC homologs, ColD (47% identity, 64% similarity) and ColE (44% identity, 60% similarity), and each CylC homolog is likely responsible for the chlorination at different positions on the alkyl chain.

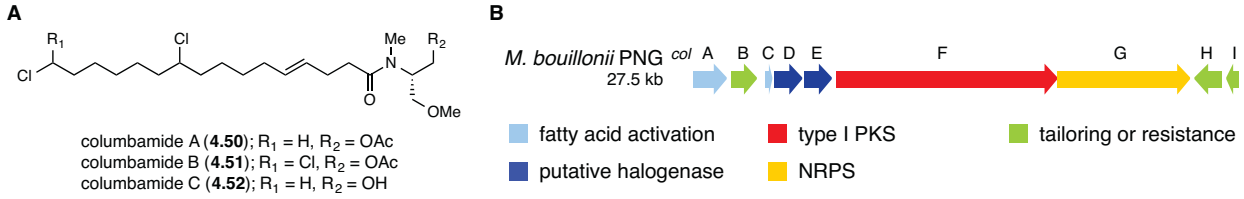
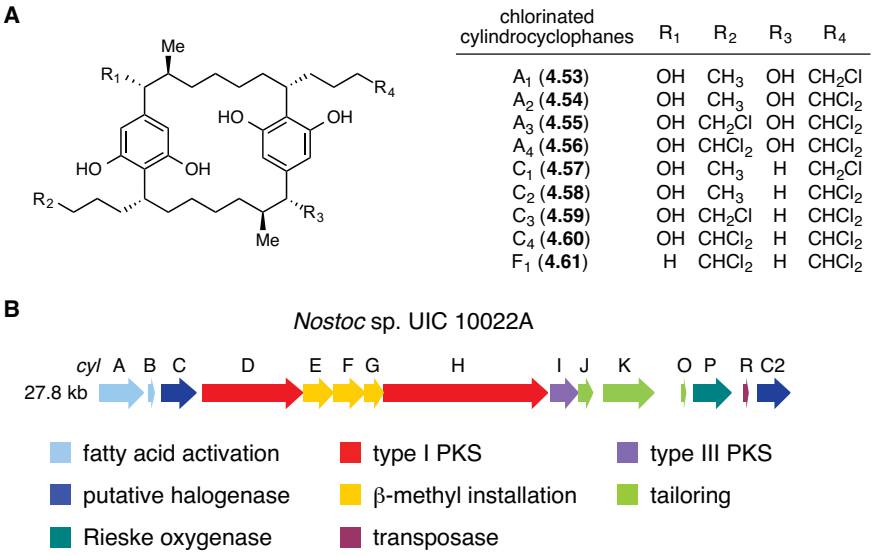


Figure 4.7: The columbamides are chlorinated natural products. **A**) Structures of the columbamides. **B**) The colubamide (*col*) biosynthetic gene cluster contains CylC homologs, ColE and ColD.

Another evidence supporting the putative halogenase function of CylC and its homologs comes from additional discoveries related to cylindrocyclophane biosynthesis. Recently, the Orjala group sequenced the genome of *Nostoc* sp. UIC 10022A, the cyanobacterial strain that produces chlorinated

cylindrocyclophanes (**4.53-4.61**, see **Figure 4.8A**).[13] The analysis of the cylindrocyclophane (*cyl*) biosynthetic gene cluster (**Figure 4.8B**) in this strain revealed an additional CylC homolog, CylC2 (38% identity, 56% similarity). We hypothesize that CylC2 is necessary for installing the chloro groups at the terminal positions of the aliphatic chains.



chlorinated cylindrocyclophanes	R_1	R_2	R_3	R_4
A_1 (**4.53**)	OH	CH_3	OH	CH_2Cl
A_2 (**4.54**)	OH	CH_3	OH	$CHCl_2$
A_3 (**4.55**)	OH	CH_2Cl	OH	$CHCl_2$
A_4 (**4.56**)	OH	$CHCl_2$	OH	$CHCl_2$
C_1 (**4.57**)	OH	CH_3	H	CH_2Cl
C_2 (**4.58**)	OH	CH_3	H	$CHCl_2$
C_3 (**4.59**)	OH	CH_2Cl	H	$CHCl_2$
C_4 (**4.60**)	OH	$CHCl_2$	H	$CHCl_2$
F_1 (**4.61**)	H	$CHCl_2$	H	$CHCl_2$

Figure 4.8: The biosynthetic gene cluster of chlorinated cylindrocyclophanes contains two CylC homologs. **A**) Structures of the chlorinated cylindrocyclophanes. **B**) The *cyl* gene cluster in *Nostoc* sp. UIC 10022A, a strain that produces the chlorinated cylindrocyclophanes.

Additionally, the Mundt group has isolated new chlorinated alkylresorcinols, cylindrofridins A-C (**4.62-4.64**, see **Figure 4.9**), from the cylindrocyclophane producing cyanobacterium, *Cylindrospermum stagnale* PCC 7417.[38] The cylindrofridins (**4.62-4.64**) are metabolites that resemble possible cylindrocyclophane biosynthetic pathway intermediates. The chlorination observed at the position on the aliphatic chain in the cylindrofridins where the C–C bond formation is expected to occur, suggesting that the cylindrocyclophane biosynthesis could proceed through cryptic chlorination.

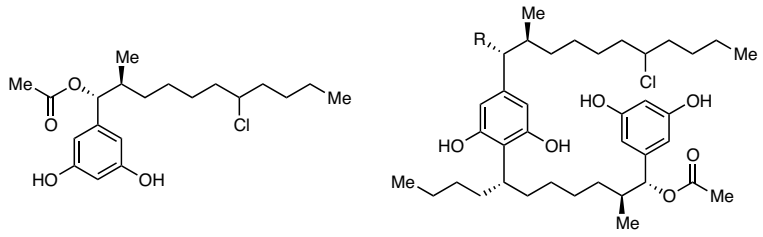


Figure 4.9: Structures of the cylindrofridins resemble the possible intermediates in the cylindrocyclophane biosynthesis if the pathway proceeds through cryptic chlorination.

If the cylindrocyclophane biosynthesis involves cryptic chlorination, then the timing of halogenation must to be considered. We hypothesized that the CylC-catalyzed chlorination could occur after the decanoic acid activation step, and CylC likely uses an ACP-bound decanoyl-CylB (**4.9**) as substrate. CylC homologs are found adjacent to a fatty acid activating enzyme (CylA homolog) and an ACP (CylB homolog) in multiple biosynthetic gene clusters, including the *mic* gene cluster (**Figure 4.5A**) and the *col* gene cluster (**Figure 4.7B**). Hence, we predictd that CylC might work together with the fatty acid activating enzymes, CylA and CylB, to catalyze chlorination early in the cylindrocyclophane biosynthetic pathway. Alternatively, CylC could act on the intermediates of the type I PKS assembly line or on the resorcinol monomer (**4.28**).

4.2.3. Purification and metal analysis of BrtJ and CylC

To study the function of CylC and its homolog from the bartoloside biosynthesis (BrtJ), the CylC homolog involved in the bartoloside B-D (**4.47**-**4.49**) biosynthesis, both enzymes were cloned into *E. coli* strains for heterologous expression. After screening various expression and purification conditions (**Table 4.1**), we succeeded in isolating N-Strep tagged BrtJ (**Figure 4.10A**). The purification of BrtJ required addition of detergent in the lysis buffer, which suggest that BrtJ may be membrane associated. In fact, the pI prediction of CylC, BrtJ and other homologs showed that these enzymes have relatively high pI values (7.0-9.5), and proteins with high pI values have been implicated to interact with the negatively charged membrane surfaces.[39] While BrtJ was solubilized in presence of detergents, the purified enzyme appears to be mostly a soluble aggregate by FPLC gel filtration analysis (**Figure 4.10B**).

Table 4.1: Expression and purification conditions screened for CylC and BrtJ (the final condition used for the purification of BrtJ is bolded)

Affinity tag	N-His_6, C-His_6, **N-Strep**, C-Strep, GST, MBP
Expression *E. coli* strain	BL21, Tuner, **Rosetta**
Expression temperature	15 ºC, **25 ºC**, 37 ºC
Expression time	6–21 h, **19 h**
Growth media	**LB**, terrific broth **with** or without **100 μM** $(NH_4)_2Fe(SO_4)_2$, **ammonium iron citrate and**/or **$MnCl_2$**
Lysis buffer	Phosphate, **EPPS**, HEPES or Tris at pH 6–9, **pH 8**
Detergent in lysis buffer	**1% Triton X-100**, 1% DDM, **8 mM CHAPS** or 20 mM Hecameg,

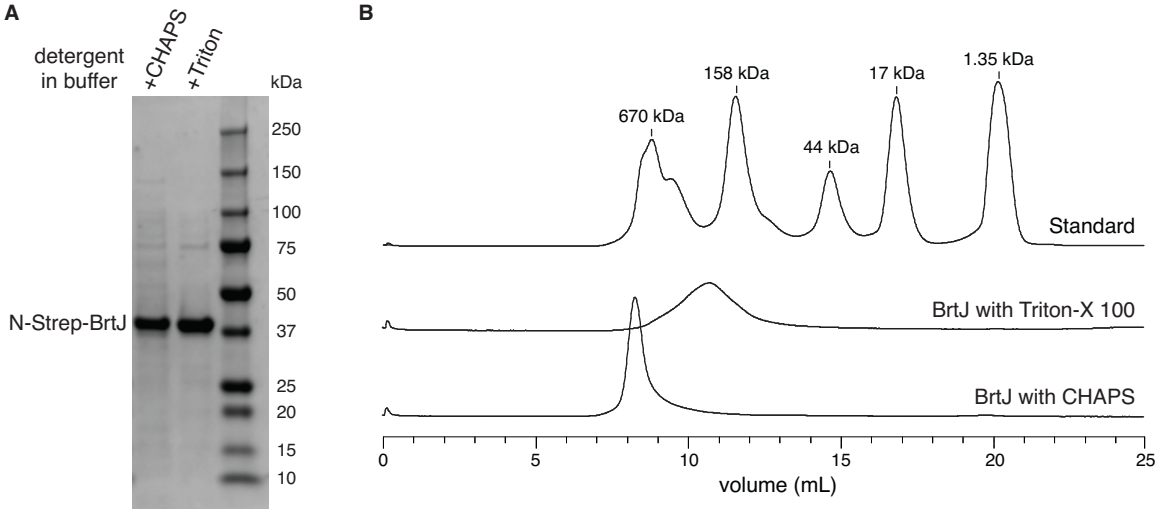


Figure 4.10: Purification of N-Strep-BrtJ. **A**) SDS-PAGE of purified N-Strep tagged BrtJ. **B**) Determination of the molecular mass of N-Strep-BrtJ by FPLC size exclusion chromatography.

Unlike BrtJ, our efforts to obtain soluble CylC through heterologous expression of CylC by itself in *E. coli* were not fruitful. Next, we tested whether the CylC expression is improved when CylA and CylB are co-expressed in *E. coli* BAP1 strain,[40] which carries a chromosomal copy of the promiscuous ppant transferase, Sfp.[41] The co-expression of CylA-C greatly enhanced CylC expression and solubility, allowing us to purify N-His_6-CylC in complex with N-His_6-CylB and a small quantity of CylA (**Figure**

4.11A). Intriguingly, we observed co-purification of untagged CylC during the purification of N-His_6-CylB through Ni-NTA affinity chromatography (**Figure 4.11B**), which indicates that the interaction between CylB and CylC is fairly strong. We also saw co-elution of CylA-C observed in the FPLC gel filtration analysis, suggesting that these three enzymes form a complex in solution (**Figure 4.11C**).

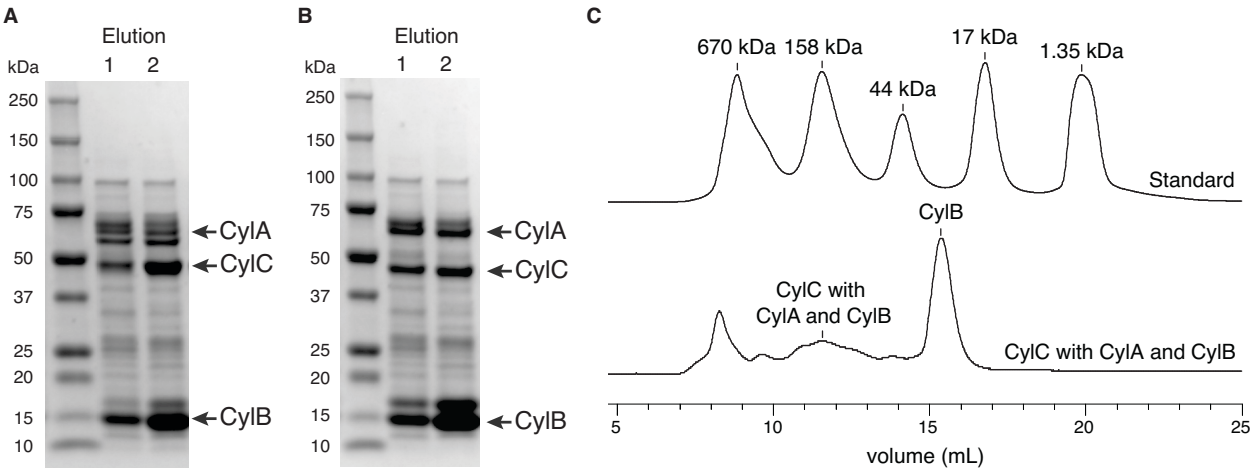


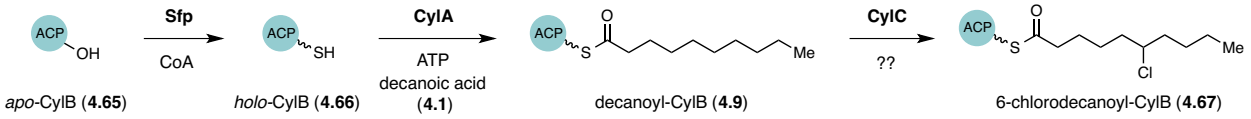
Figure 4.11: Co-expression of CylA, CylB and CylC. **A**) SDS-PAGE of purified N-His_6-CylC and N-His_6-CylB co-eluting with small amount of CylA. **B**) SDS-PAGE of untagged CylA and CylC co-eluting with N-His_6-CylB. **C**) FPLC size exclusion chromatography of the co-purified CylA, CylB and CylC.

The amount of iron in the N-Strep-BrtJ and the N-His_6-CylC obtained through heterologous expression was quantified using the Ferene S spectrophotometric assay.[42] The assay result indicates that there is approximately one equivalent of iron found in the BrtJ sample (0.98 mol of iron per 1 mol of BrtJ). In addition, the quantity of iron in the samples of containing CylA, CylB and CylC appeared to positively correlate with the amount of CylC in the protein samples (16.8 μM of iron in 1 mg/mL of protein solution containing mostly N-His_6-CylC and N-His_6-CylB, see **Figure 4.11A**; 7.7 μM of iron in 1 mg/mL of protein for sample in the gel **Figure 4.11B**). No iron was detected in a protein sample only containing N-His_6-CylB, which was purified from the negative control BAP1 *E. coli* strain co-expressing CylA and CylB, but not CylC. This result support our hypothesis that BrtJ, CylC and their homologs are iron-dependent enzymes, although we cannot rule out the possibility that other metals, such as manganese, may also be required for catalysis. Despite much effort, including anaerobic reconstitution using

ammonium iron sulfate and various reducing agents (sodium dithionite, flavodoxin/flavodoxin reductase, phenazine methosulfate/NADH, and spinach ferredoxin/ferredoxin reductase), we have not yet been able to observe activities of the purified BrtJ or CylC *in vitro*. Reconstitution and biochemical characterizations of BrtJ and CylC are actively being pursued.

4.2.4. *In vivo* CylC activity assay in the *E. coli* co-expression strain

Since the *in vitro* reconstitution of CylC has been challenging, we decided to focus on analyzing the activity of CylC *in vivo*. The predicted substrate for CylC, decanoyl-CylB (**4.9**, see **Scheme 4.9**), was generated *in situ* by feeding decanoic acid to the *E. coli* BAP1 culture co-expressing CylA-CylC. The ppant transferase Sfp converts the ACP CylB from the *apo* form (**4.65**) to the active *holo* form (**4.66**), and CylA is expected to activate decanoic acid (**4.1**) for loading onto the *holo*-CylB (**4.66**).



Scheme 4.9: *In vivo* CylC activity assay in *E. coli* BAP1 co-expressing CylA, CylB and CylC.

N-His_6-CylB was then purified from the co-expression strain and trypsin digested. The digested CylB peptide containing the active site serine residue with the ppant modification (**4.68**, see **Figure 4.12A**) was detected by LC-HRMS. The detected peptide with the ppant modification was further analyzed by MS/MS experiments to observe the masses corresponding to the acyl moiety (**4.69**) attached to the ppant arm. The masses corresponding to the ppant-modified peptide of Cl-decanoyl-CylB (**4.70**, see **Figure 4.12B**) and the Cl-decanoyl moiety (**4.71**) were observed in the LC-HRMS and MS/MS spectra, respectively (**Figure 4.12C**). These masses were absent in the negative control sample of N-His_6-CylB purified from the *E. coli* BAP1 lacking CylC.

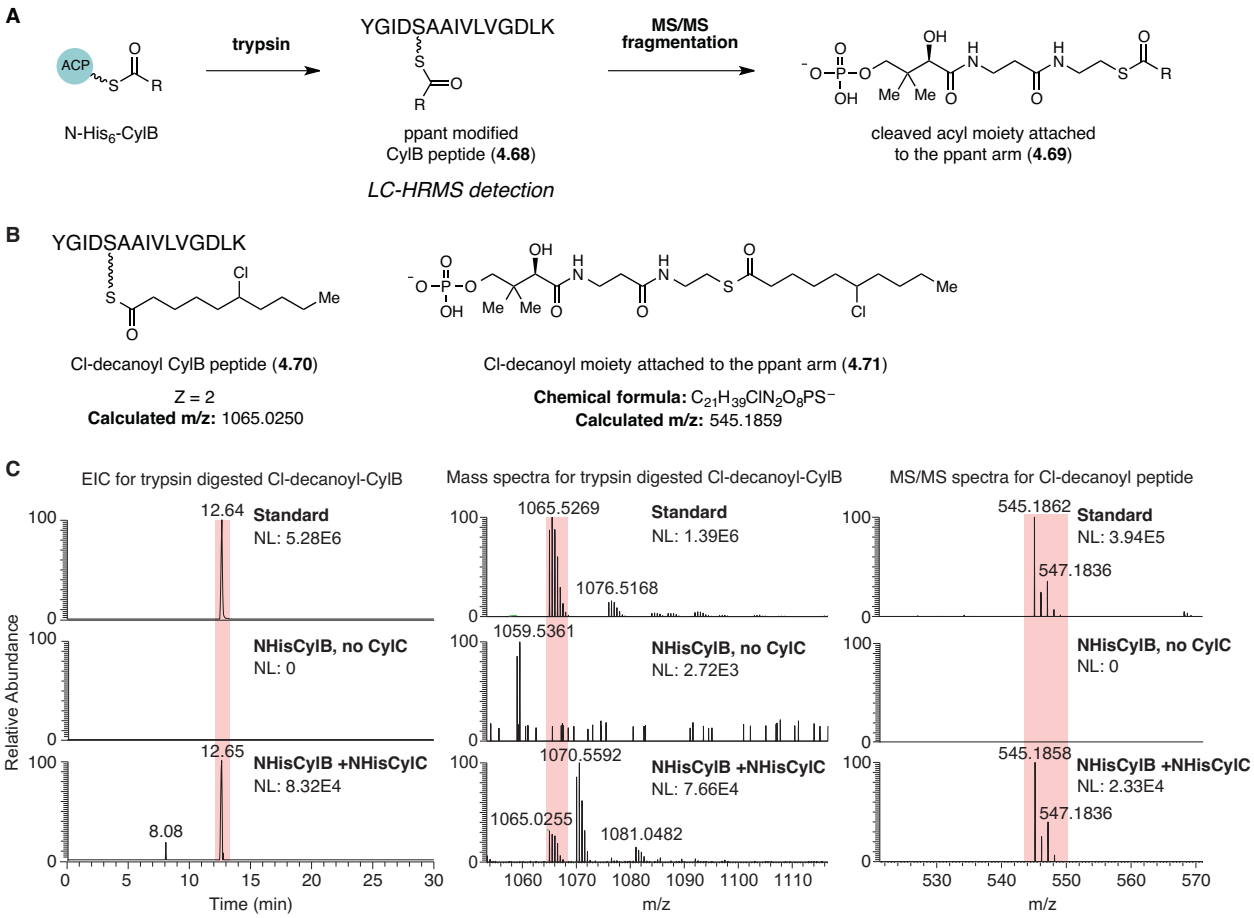
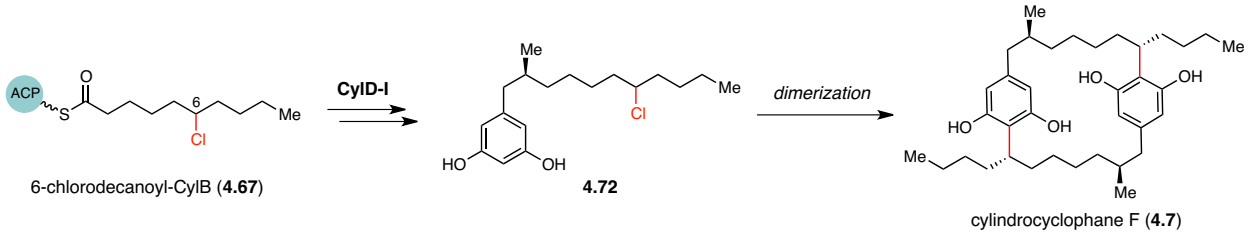


Figure 4.12: Detection of chlorinated decanoyl-CylB peptide and the chlorinated decanoyl moiety by LC-HRMS and MS/MS. **A**) LC-HRMS and MS/MS analysis of N-His$_6$-CylB purified from the co-expression strain. **B**) The structures and the calculated m/z values of the target peptide and the acyl moiety. **C**) Observation of the Cl-decanoyl-CylB peptide and the Cl-decanoyl moiety in the N-His$_6$-CylB sample purified from the *E. coli* BAP1 strain co-expressing CylA and CylC.

The detection of the chlorinated decanoyl-CylB (**4.67**) in the *in vivo* CylC activity assay is the first experimental verification of the halogenase function of CylC. Halogenase activity has never been reported for the di-metallo carboxylate enzymes. While we have not yet confirmed the exact position of the chlorine atom on the acyl chain of decanoyl-CylB observed in this assay, the chlorination is expected to occur at C6 position to prime the molecule for subsequent C–C bond formation (**Scheme 4.10**). The investigation of this C–C bond formation will be discussed in **Chapter 5**. Further studies, such as the use of the decanoic acid substrate with deuterium-labels at C6, will be done to determine the site of CylC-

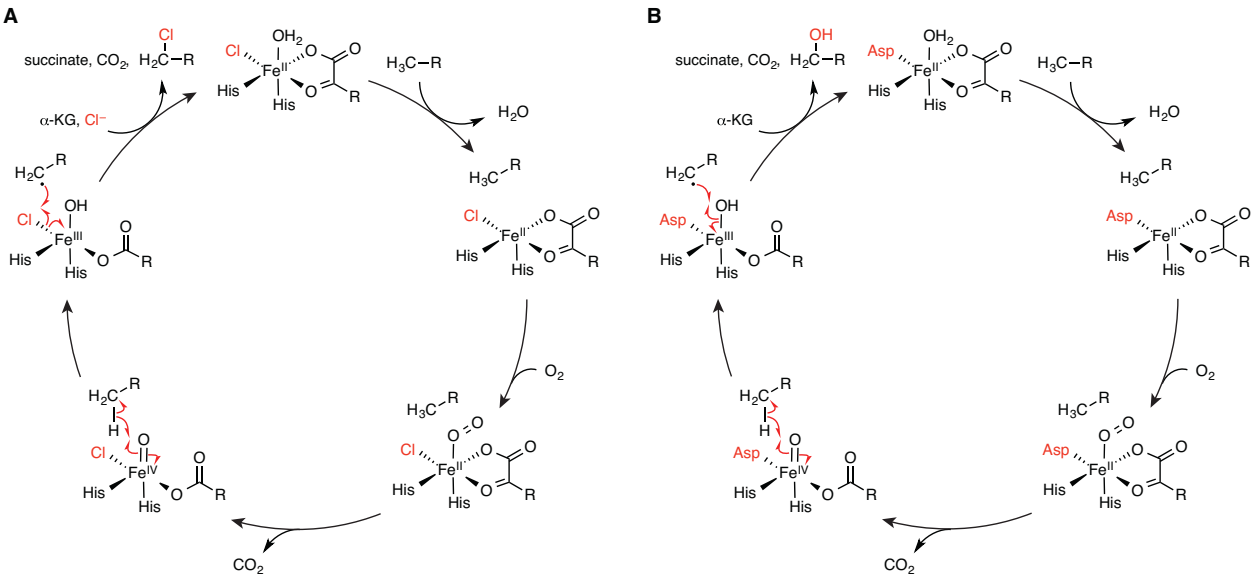
catalyzed chlorination. The halogenase activity of CylC observed in this *in vivo* assay support our hypothesis that the cylindrocyclophane biosynthesis involves cryptic chlorination.



Scheme 4.10: Chlorine substituent on the Cl-decanoyl-CylB is expected to be on C6 for the subsequent C–C bond formations in the dimerization step.

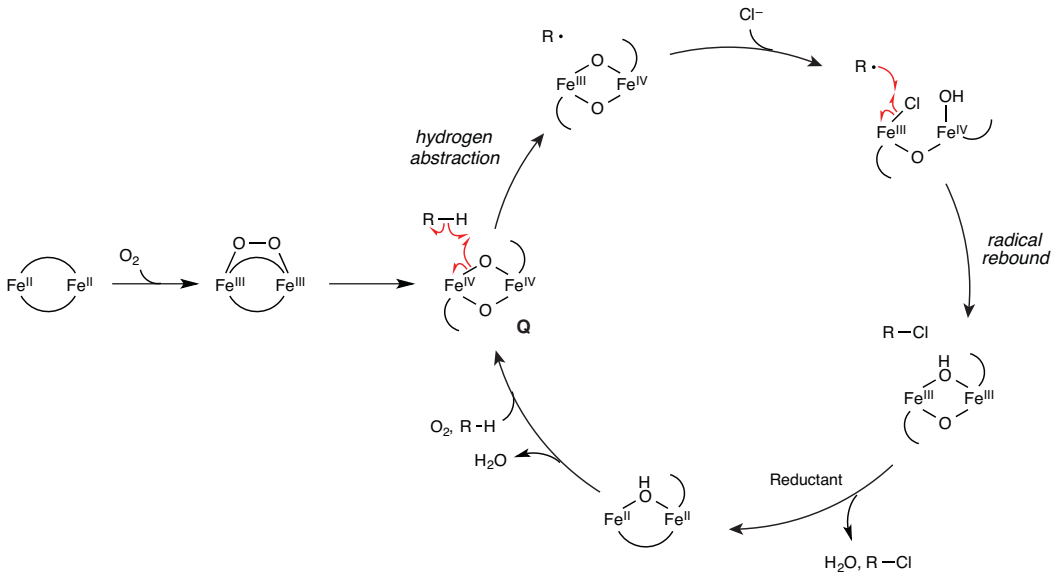
4.2.5. Proposed mechanism and distribution of the di-metallo carboxylate halogenases

The type of halogenation catalyzed by CylC and its homologs resembles the radical halogenation catalyzed by the known □-ketoglutarate non-heme iron dependent halogenases.[43] The mechanism of chlorination catalyzed by the □-ketoglutarate non-heme iron dependent halogenases involves a radical abstraction of the hydrogen to form an alkyl radical (**Scheme 4.11A**). Then, a radical rebound with a chloride results in the alkyl chloride formation.[43] This mechanism is very similar to that of the □-ketoglutarate non-heme iron dependent hydroxylases, in which the final radical rebound occurs with a hydroxyl group instead of a chloride to form an alcohol product (**Scheme 4.11B**).[44] One important difference between the □-ketoglutarate non-heme iron dependent halogenases and hydroxylases is the active site residues involved in the iron binding. The hydroxylases use two histidine and one carboxylate (Asp or Glu) residues, which for a facial triad to bind iron.[45] On the other hand, the active site iron in the halogenases is coordinated by only two histidine residues, and the third carboxylate residue is replaced by a chloride. The loss of the carboxylate ligand and its replacement by chloride is essential for the halogenation activity of the □-ketoglutarate non-heme iron dependent halogenases.[46]



Scheme 4.11: Mechanisms of the α-ketoglutarate non-heme iron dependent halogenase and hydroxylase.[43] **A**) Halogenation involves a radical rebound with a chloride ligand. **B**) Hydroxylases have a 2-His-1-Asp/Glu facial triad, and the mechanism involves a radical rebound with a hydroxyl ligand.

As in the cases of the α-ketoglutarate non-heme iron dependent halogenases, we hypothesize that chlorination of unactivated aliphatic carbon centers catalyzed by CylC and its homologs proceed through a radical mechanism that involves a hydrogen abstraction by the di-metallo cofactor followed by a radical rebound with a chloride (**Scheme 4.12**). We predict that the chloride that participates in the final radical rebound is coordinated to the di-iron (or other metals such as manganese) cofactor. The chlorination mechanism shown in **Scheme 4.12** is adopted from the hydroxylation mechanism of the known di-iron carboxylate enzymes, such as the soluble methane monooxygenase.[28,47] The reaction initiates with the activation of dioxygen by the reduced di-iron cofactor, leading to the hydrogen abstraction by the oxidized intermediate Q. The resulting alkyl radical could react with the chloride ligand that coordinates di-iron cofactor to form the alkyl chloride product. Further work on CylC and its homologs are necessary to determine the true identity of the metal cofactors and the mechanism of halogenation.



Scheme 4.12: Proposed mechanism of the chlorination catalyzed by CylC.

We performed bioinformatic searches to determine the abundance of CylC homologs in comparison to the other types of halogenases. Because CylC and its homologs appear to be present only in cyanobacteria, we limited our searches to cyanobacterial genomes. We carried out BLAST searches to look for the homologs of the flavin-dependent halogenase RebH,[48] the □-KG dependent non-heme iron halogenase SyrB2,[49] and the □-KG dependent non-heme iron halogenases WelO5 that does not require an ACP-bound substrate.[50] Through this search, we identified 18 RebH homologs, 19 SyrB2 homologs, and 28 WelO5 homologs in the genomes of cyanobacteria available in the NCBI database (E-value cut-off of <E-20).

The BLAST search using CylC as a query returned 67 sequences that had high homology to the full sequence of CylC (coverage of >70%). The multiple sequence alignment of the hits revealed that all seven putative metal-binding residues (3 His and 4 Glu) are conserved in 34 sequences (**Table 4.2**), while the other 33 sequences are missing one of the glutamate residues (Glu-283). To use the most conservative estimate, we only considered the 34 sequences that had all seven putative metal-binding residues as true CylC homologs (E value < E-65). Based on these BLAST searches, the putative dimetallo carboxylate halogenases (CylC homologs) appeared to be a major type of halogenases in cyanobacteria (**Figure 4.13**).

Table 4.2: The CylC homologs with the conserved metal-binding residues.

Protein	Accession	Organism	Length	Homology (%ID, %similarity)
CylC	AFV96137.1	*Cylindrospermum licheniforme* UTEX B2014	471	100/100
CylC homolog	WP_015207402.1	*Cylindrospermum stagnale* PCC 7417	471	98/99
CabC	AMB48447.1	*Nostoc* sp. CAVN2	471	96/97
Unknown	WP_036264917.1	*Mastigocoleus testarum*	467	52/70
Unknown	WP_051470326.1	*Fischerella* sp. PCC 9605	453	53/68
Unknown	WP_019489578.1	*Calothrix* sp. PCC 7103	456	52/67
Mic homolog	WP_002752271.1	*Microcystis aeruginosa*	465	53/68
Unknown	WP_013334462.1	*Cyanothece* sp. PCC 7822	454	51/69
Unknown	WP_026798333.1	*Planktothrix prolifica*	465	51/68
Unknown	WP_015142039.1	*Pleurocapsa minor*	473	48/63
Unknown	WP_002804147.1	*Microcystis aeruginosa* PCC 9701	458	52/67
Unknown	WP_016949101.1	*Anabaena* sp. PCC 7108	459	49/65
Unknown	WP_016871427.1	*Fischerella thermalis*	408	51/63
Unknown	WP_009459338.1	*Fischerella* sp. JSC-11	408	51/63
Unknown	WP_012409761.1	*Nostoc punctiforme* PCC 73102	456	52/66
ColD	AKQ09581.1	*Moorea bouillonii* PNG5-198	463	47/65
Unknown	WP_006527875.1	*Gloeocapsa* sp. PCC 73106	464	48/63
Unknown	WP_053455402.1	*Hapalosiphon* sp. MRB220	410	51/63
Unknown	WP_026719275.1	*Fischerella* sp. PCC 9431	410	50/62
Unknown	WP_045054787.1	Chroococcales cyanobacterium CENA595	470	47/65
ColE	AKQ09582.1	*Moorea bouillonii* PNG5-198	470	47/62
Unknown	WP_026724087.1	*Fischerella* sp. PCC 9431	468	47/61
Unknown	WP_053457689.1	*Hapalosiphon* sp. MRB220	468	46/61
Unknown	WP_053457697.1	*Hapalosiphon* sp. MRB220	457	45/62
Unknown	KOP24665.1	*Hapalosiphon* sp. MRB220	452	45/62
Unknown	WP_051206793.1	*Fischerella* sp. PCC 9431	457	45/62
Unknown	WP_019504306.1	*Pleurocapsa* sp. PCC 7319	446	45/62
Unknown	WP_015328291.1	*Cylindrospermum stagnale* PCC 7417	456	44/61

(Continued)

Unknown	BAU67476.1	*Stanieria* sp. NIES-3757	451	40/58
Unknown	WP_016949100.1	*Anabaena* sp. PCC 7108	442	44/58
Unknown	AFZ58482.1	*Anabaena cylindrica* PCC 7122	403	42/56
Unknown	WP_042466014.1	*Anabaena cylindrica*	400	42/56
Unknown	WP_016870033.1	*Fischerella muscicola*	410	53/69
Unknown	WP_026087261.1	*Fischerella muscicola*	405	51/66
BrtJ	AKV71855.1	*Synechocystis salina* LEGE 06155	401	49/66

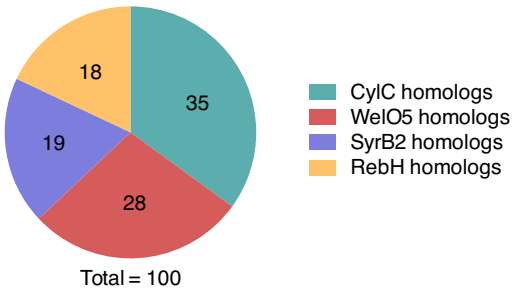


Figure 4.13: Distribution of the four types of halogenases shows that CylC homologs are a major type of halogenase in cyanobacteria.

4.2.6. CylH and CylI assays with chlorinated substrates

The *in vivo* CylC activity assay result suggests that the chlorination step likely occurs early in the cylindrocyclophane biosynthetic pathway. *In vitro* CylA substrate scope assay using unfunctionalized (**4.1**) and pre-functionalized substrates (**4.73** and **4.74**, see **Figure 4.14A**) showed that CylA only activates the unfunctionalized decanoic acid (**Figure 4.14B**), confirming our hypothesis that CylC acts after decanoyl-CylB (**4.9**) formation.

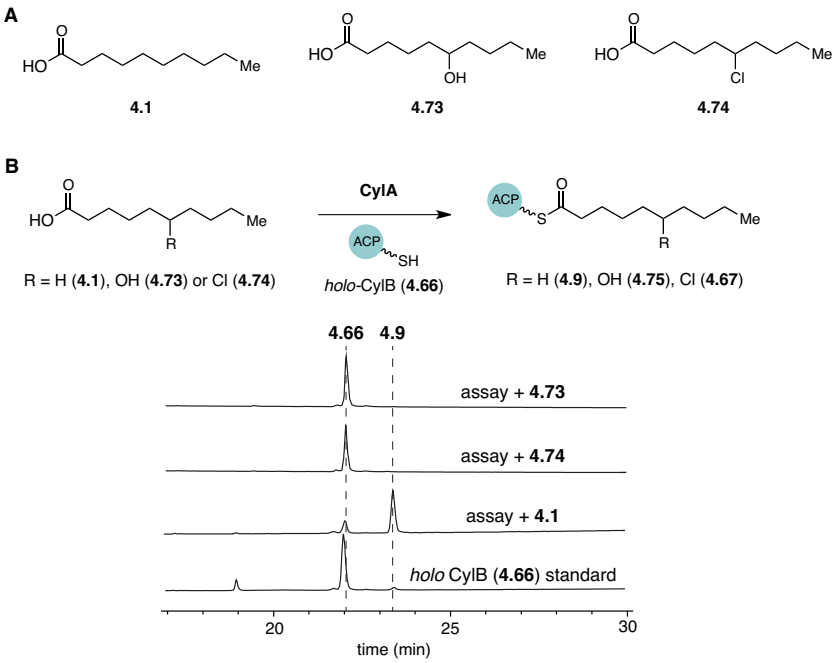
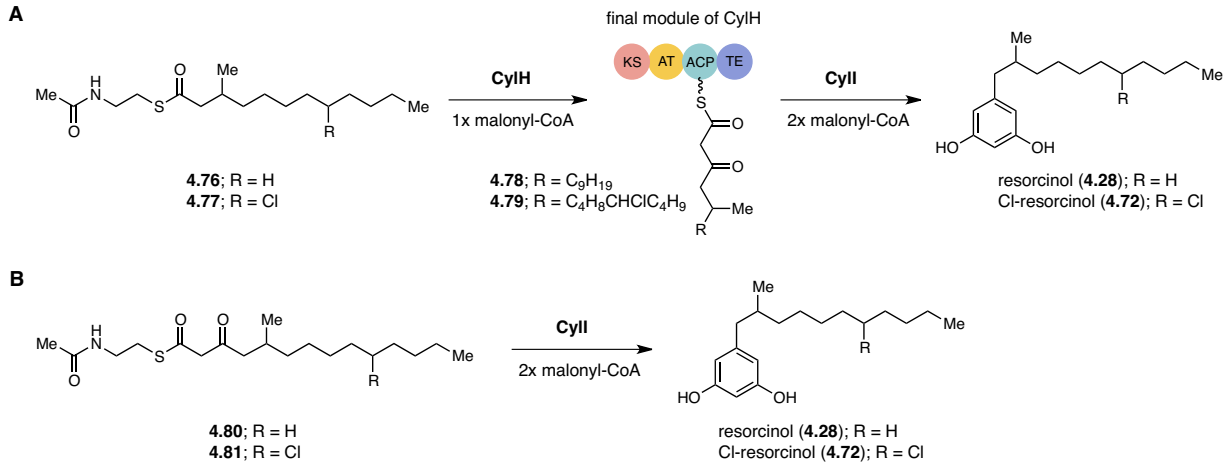


Figure 4.14: CylA substrate scope assays. **A**) Decanoic acid derivatives tested in CylA activity assays. **B**) HPLC CylA assays testing the loading of decanoic acid derivatives onto *holo* CylB (monitored at 220 nm).

All of our previous *in vitro* characterization of the other enzymes in the pathway (see **Chapters 2** and **3**),[1,51] including studies of the final module of the type I PKS CylH and the type III PKS CylI, were performed with non-chlorinated substrates (**4.76** and **4.80**, see **Scheme 4.13**). If the CylC-catalyzed chlorination takes place after decanoic acid activation by CylA and CylB, then the downstream enzymes must process chlorinated substrates. To test whether this was the case, chlorinated SNAC substrates (**4.77** and **4.81**) for CylH and CylI were synthesized to assess the substrate scope of these enzymes *in vitro*. The HPLC analysis of the assay mixtures (**Figure 4.15**) showed that both CylH and CylI are capable of using chlorinated substrates to form a chlorinated resorcinol product (**4.72**).



Scheme 4.13: *In vitro* assays of CylI and the final module of CylH using non-chlorinated or chlorinated substrates. **A**) Coupled CylH and CylI assays. **B**) CylI assay.

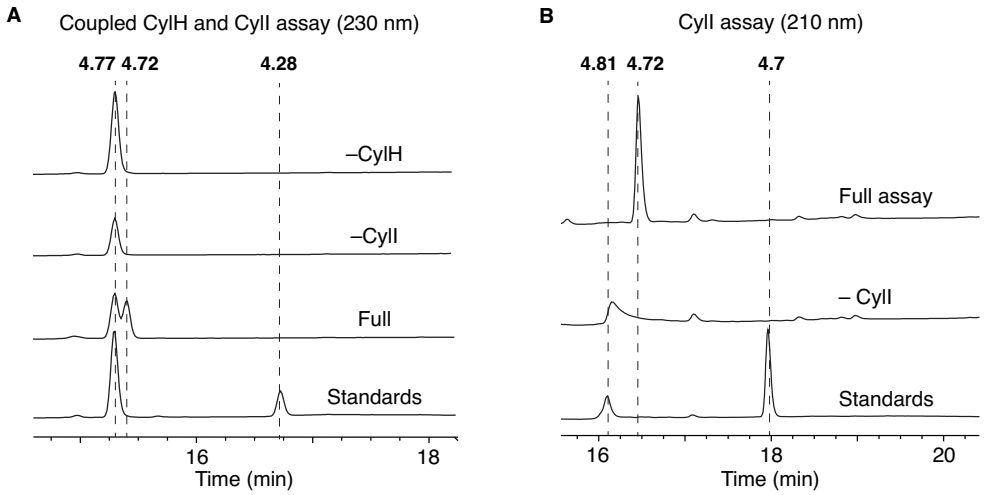


Figure 4.15: HPLC analysis of the CylH and CylI assays using chlorinated substrates. **A**) Coupled assay with the final module of the type I PKS CylH and the type III PKS CylI using **4.77**. **B**) Assay with CylI using **4.81**. Standard contains substrate **4.81** and cylindrocyclophane F **4.7** (shift in retention time is observed due to an addition of a guard column).

We performed a time course analysis with non-chlorinated and chlorinated substrates to determine which substrates to determine the preferred substrates for CylH and CylI. The time course analysis of the CylI assay using non-chlorinated (**4.80**) and chlorinated substrates (**4.81**) revealed that the rate of turnover appears comparable for both substrates (**Figure 4.16A**). In addition, the qualitative conversion in the coupled CylH and CylI assays at two-hour time point also appear similar for both non-chlorinated (**4.76**) and chlorinated substrates (**4.77**, **Figure 4.16B**).

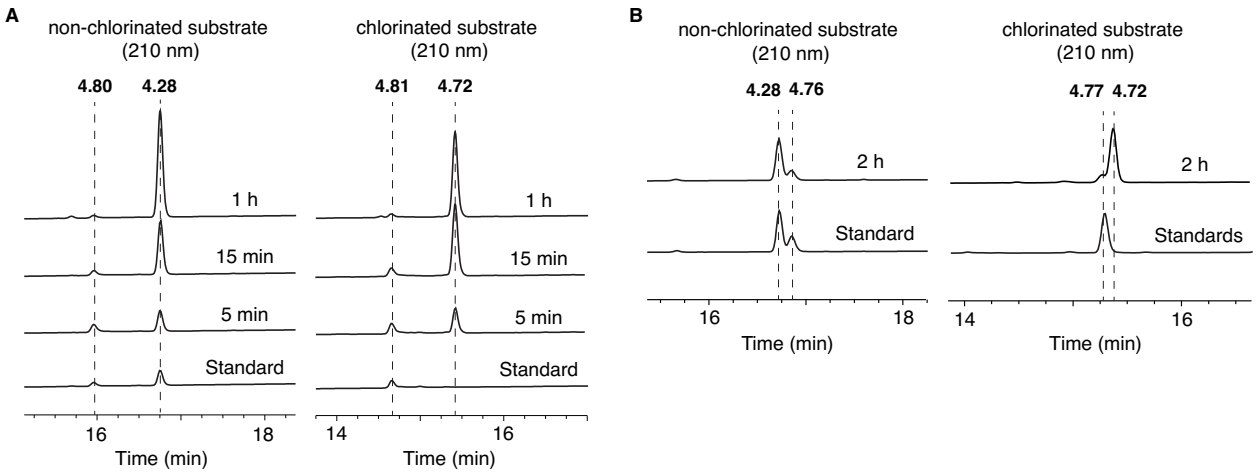


Figure 4.16: Comparison of the CylH and CylI assays performed with non-chlorinated or chlorinated substrates. **A**) Time course analysis of the CylI assay using the □-ketoacyl-SNAC substrates (**4.80** or **4.81**). **B**) Comparison of the two-hour time point of the coupled CylH and CylI assays using acyl-SNAC substrates (**4.76** or **4.77**).

Analysis of the substrate scope of CylH and CylI showed that these two enzymes have some degree of promiscuity and accept both non-chlorinated and chlorinated substrates to similar degrees. These results indicate that there is a possibility that the chlorination occurs before the processing by the type I PKS enzymes in cylindrocyclophane biosynthesis. Another important finding is that the *in situ* generation of chlorinated resorcinol (**4.72**) did not result in spontaneous dimerization to form cylindrocyclophane F (**4.7**). The lack of spontaneous paracyclophane formatio in these assays suggests that a separate enzyme is required for forming the C–C bonds to construct the macrocycle.

4.3. Conclusions

Through our assessment of the two potential routes for the paracyclophane formation, we determined the functions of two metalloenzymes, CylP and CylC, which were predicted to have the ability to functionalize an unactivated carbon center. The Rieske oxygenase CylP is responsible for the benzylic oxidation of monomeric resorcinol (**4.28**) to form hydroxylated resorcinol (**4.29**). The results of our assays using the CylP-containing lysates indicate that the benzylic hydroxylation occurs prior to the dimerization step. The benzylic hydroxylation of resorcinol (**4.28**) can lead to the production of cylindrocyclophanes A-E (**4.2**-**4.6**). The observed activity of CylP therefore did not support its role in

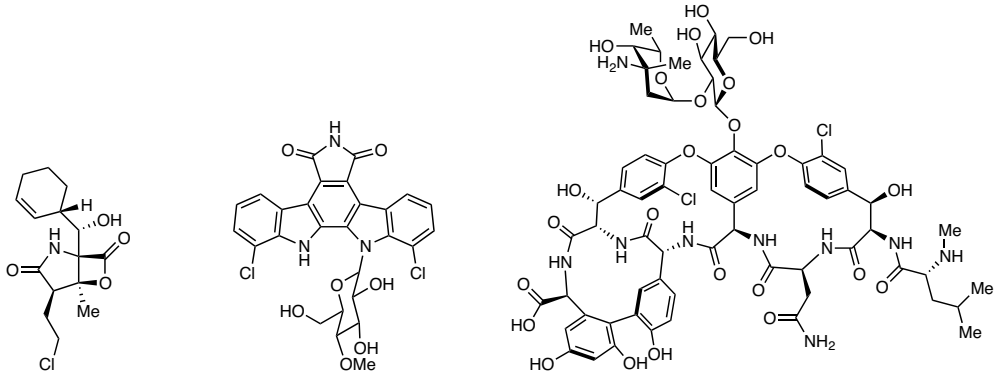
paracyclophane formation via the direct alkylation route, suggesting that the cylindrocyclophane biosynthesis proceeded through an alternative route.

The other candidate enzyme, CylC, was identified through bioinformatics to be a putative halogenase belonging to the family of di-iron carboxylate enzymes through bioinformatics. We predicted that the CylC-catalyzed chlorination occured after decanoic acid activation by CylA and CylB, and the observed interaction between CylB and CylC during purification supported our hypothesis. Indeed, we succeeded in demonstrating the halogenase function of CylC for the first time through an *in vivo* activity assay. The observed formation of chlorinated decanoyl-CylB (**4.67**) suggested that the cylindrocyclophane biosynthesis proceeds through the pre-functionalization route involving cryptic chlorination. In addition, the substrate scope of CylA, CylH and CylI also confirm the predicted early timing of the chlorination step.

Although we have been able to verify the halogenase function of CylC through an *in vivo* assay, there are many unanswered questions that require further investigations. While we have observed the mass corresponding to chlorinated decanoyl-CylB (**4.67**) using LC-HRMS, we have not yet verified the exact location of the chlorine substituent on the acyl chain. Additional *in vivo* assays using our current co-expression strain using isotopically labeled decanoic acids are needed to confirm the predicted regio- and stereoselectivity of CylC. Another question to be answered is the promiscuity of CylC toward different halide ions. Previously characterized α-ketoglutarate non-heme dependent halogenases, flavin-dependent halogenases, and SAM-dependent halogenases are known to be relatively promiscuous and use different halides (Cl^-, Br^-, I^-).[52,53] The isolations of the brominated cylindrocyclophanes[13] and carbamidocyclophanes[54] from the KBr-enriched cultures of *Nostoc* sp. UIC 10022A and *Nostoc* sp. CAVN2, respectively, suggest that CylC and its homologs are capable of catalyzing bromination as well. More mechanistic work of CylC and its homologs requires *in vitro* reconstitution and crystal structures of these enzymes. Currently, an effort to crystallize BrtJ and CylC is ongoing in collaboration with the Drennan group at the Massachusetts Institute of Technology.

Our bioinformatic search indicates that CylC homologs are widely distributed among cyanobacteria, some of which are found in the biosynthetic pathways of previously reported chlorinated natural products.[36,37] Intriguingly, CylC homologs are only found in cyanobacteria and appear to be absent in other bacterial phyla. Cyanobacteria are known to be prolific producers of secondary metabolites, many of which are halogenated.[31] The presence of CylC homologs in many of the sequenced cyanobacterial genomes indicate that these novel halogenases play a significant role in the production of halogenated metabolites in cyanobacteria. Moving forward, the distribution of CylC homologs and other known classes of halogenases in cyanobacteria will be investigated.

To conclude, more than 4000 halogenated natural products have been discovered, and approximately half of those are organochlorines.[55] Chlorinated natural products include potent anticancer agents such as β-lactone salinisporamide A (**4.82**, see **Figure 4.17**) and rebeccamycin (**4.83**) as well as antibiotics such as vancomycin (**4.84**), and the presence of chlorine substituents in these molecules is important for their biological activities.[46,56-58] Enzymatic halogenations, especially those involving functionalization of the unactivated carbon centers, are remarkable from both biological and chemical perspectives. Our effort to understand the biosynthesis of the cylindrocyclophanes with an atypical paracyclophane scaffold led to the discovery of the novel di-metallo carboxylate halogenases, an important addition to the previously known classes of halogenases.[43]



salinisporamide A (**4.82**) *anticancer*

rebeccamycin (**4.83**) *anticancer and antimicrobial*

vancomycin (**4.84**) *antibiotic*

Figure 4.17: Chlorinated natural products for which the presence of chlorine substituent has been demonstrated to affect their bioactivity.

4.4. Materials and Methods

4.4.1. Materials and General Methods

All chemical reagents were purchased from Sigma–Aldrich, unless otherwise indicated. Luria-Bertani Lennox (LB) medium was purchased from EMD Millipore or Alfa Aesar and IPTG was purchased from Teknova. All NMR solvents were purchased from Cambridge Isotope Laboratories. Primers were purchased from Integrated DNA Technologies. Thermocycling was carried out in a C1000 Gradient Cycler (Bio-Rad). DNA polymerases, restriction enzymes and T4 ligase for cloning were purchased from New England BioLabs. Recombinant plasmid DNA was purified with a QIAprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean up were performed using an Illustra GFX PCR DNA and Gel Band Purification kit from GE Healthcare. DNA sequencing was performed by Beckman Coulter Genomics (Danvers, MA). Nickel-nitrotriacetic acid-agarose (Ni-NTA) resin was purchased from Qiagen and Thermo Scientific, and *Strep*-Tactin resin was purchased from IBA. SDS-PAGE gels were purchased from Bio-Rad.

DNA concentrations were determined using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific). Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm. Gel filtration FPLC was performed on a BioLogic DuoFlow Chromatography System (Bio-Rad) equipped with a Superdex column, 200 10/300 GL (GE Healthcare). Bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa), and vitamin B_{12} (1350 Da) were used as molecular weight markers (Bio-Rad #151-190). The molecular weights of the proteins analyzed by gel filtration were calculated from their elution volume, using a linear relationship between log(molecular weight) and ve/vo (elution volume/void volume). Protein concentrations were determined using a NanoDrop 2000 UV-Vis Spectrophotometer for NStrep-BrtJ (ε = 74,510 $M^{-1}cm^{-1}$), $NHis_6$-CylB (ε = 18.450 $M^{-1}cm^{-1}$) and $NHis_6$-CylC (ε = 79,440 $M^{-1}cm^{-1}$). The extinction coefficients were calculated using ExPASy ProtParam tool.

Analytical HPLC was performed on a Dionex Ultimate 3000 instrument (Thermo Scientific). High-resolution mass spectral data and the spectrophotometric assay data were obtained in the Small Molecule Mass Spectrometry Facility, FAS Division of Science. Mass spectral data were obtained on an Agilent 1290 Infinity UHPLC system (Agilent Technologies) and a maXis impact UHR time-of-flight mass spectrometer system (Bruker Daltonics Inc.) equipped with an electrospray ionization (ESI) source. High-resolution mass spectral (HRMS) data for the synthetic compounds were obtained on an Agilent 6210 TOF. The capillary voltage was set to 4.5 kV and the end plate offset to –500 V, the drying gas temperature was maintained at 190 ^{o}C, with a flow rate of 8 L/min and a nebulizer pressure of 21.8 psi. The liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with 50% H_2O and 50% acetonitrile as solvent.

Proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were obtained on a Varian Inova-500 (500 MHz, 125 MHz) NMR spectrometer in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance as internal standard for ^{1}H ($CDCl_3$, δ_H = 7.26 ppm) and ^{13}C ($CDCl_3$, δ_C = 77.2 ppm). Data are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, m = multiplet), integration, and coupling constant. NMR spectra were visualized using iNMR Version 5.3.6 (Mestrelab Research).

4.4.2. Cloning of *cylP*

Table 4.3. Oligonucleotides used for cloning *cylP* (restriction sites underlined).

Primer name	Sequence (5' to 3')	Target
CylP-MBP-F	GCCGCCGC<u>GAATTC</u>ATGTACCAAATTAACC	MBP-CylP
CylP-MBP-R	GCCGCC<u>CTGCAG</u>TTAATCATTGTGAAAATG	

The PCR amplification of *cylP* from *Cylindrospermum licheniforme* ATCC 29412 genomic DNA was carried out using the primers shown in **Table 4.3**. The forward primer has an *Eco*RI restriction site and the reverse primer has a *PstI* restriction site. The PCR reaction contained Phusion High-Fidelity PCR Master Mix, 60 ng DNA template, and 25 pmoles of each primer in a total volume of 50 µL. The thermocycling was carried out using the following parameters: denaturation for 30 sec at 98 ºC, followed by 35 cycles of 10 sec at 98 ºC, 30 sec at 54 ºC, 1 min at 72 ºC, and a final extension time of 10 min at 72 ºC. PCR products were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled, and purified.

Amplified fragments were digested with *Eco*RI and *Pst*I for 2.5 h at 37 ºC. Digests contained 0.5 µL water, 5 µL NEB buffer 3.1 (10x), 20 µL of PCR product, 1.5 µL of *Pst*I (20 U/µL), and 3 µL of *Xho*I (20 U/µL). Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digests were ligated into linearized pMAL-c2X expression vector using T4 DNA ligase. Ligations were incubated at 16 ºC overnight and contained 3 µL water, 1 µL T4 Ligase Buffer (10x), 1 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA Ligase (400 U/µL). The ligation sample (5 µL) was used to transform a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pMAL-c2X-*cylP* construct was confirmed by sequencing of the purified plasmid DNA. The pMAL-c2X-*cylP* was transformed into chemically competent *E. coli* BL21 (DE3) cells (Invitrogen). The resulting transformants were stored at –80 ºC as frozen 1:1 LB/glycerol stocks.

4.4.3. Feeding experiment in CylP-expressing *E. coli* strain

A 5 mL starter culture of pMAL-c2X-*cylP* BL21 *E. coli* was inoculated with frozen stock and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL ampicillin. Overnight cultures were diluted 1:100 into 50-mL LB medium containing ampicillin. The cultures were incubated at 37 ºC with 175 rpm shaking until OD_{600} = 0.45–0.5. The cultures were then split into 5 mL aliquots and 100 µM resorcinol (**4.28**) was added as DMSO stock (100 mM). The protein expression was induced with 100 µM

of IPTG and the cultures were incubated at 25 ºC for 18 h with 200 rpm shaking. Each culture was extracted with ethyl acetate (2 × 5 mL) and the combined organic layer was concentrated *in vacuo*. The dried samples were re-dissolved in 3:1 acetonitrile/water solution containing 2% DMSO. The resuspended samples were centrifuged at 16,100 ×*g* for 10 min and the supernatants were analyzed by LC-HRMS and LC-HRMS/MS.

5 µL samples were analyzed by LC-HRMS on a Gemini C18 analytical column (4.6 x 50 mm, 5 µm) equipped with a guard column at a flow rate of 0.4 mL/min with the column temperature maintained at room temperature. The following elution conditions were applied: 75% solvent B in solvent A for 2 min, a gradient increasing from 75% to 100% solvent B in solvent A over 15 min, 100% solvent B for 3 min, a gradient decreasing to 100% to 75% solvent B in solvent A over 0.1 min, and 75% solvent B in solvent A for 10 min. (solvent A = 95:5 water/acetonitrile with 10 mM ammonium hydroxide; solvent B = 92:2 acetonitrile/water in 10 mM ammonium hydroxide). For the MS detection, the ESI mass spectra data were recorded on a negative ionization mode for a mass range of m/z 50 to 3000. A mass window of ± 0.005 Da was used to extract the ion of $[M-H]^-$. For ions of interest, MS/MS analysis was performed using fragmentation voltage of 30.0 eV.

4.4.4. *In vitro* assay with the lysate of CylP-expressing *E. coli* strain

A 5 mL starter culture of pMAL-c2X-*cylP* BL21 *E. coli* was inoculated with frozen stock and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL ampicillin. Overnight cultures were diluted 1:100 into 50 mL LB medium containing ampicillin. The cultures were incubated at 37 ºC with 175 rpm shaking until OD_{600} = 0.35–0.4. The protein expression was induced with 100 µM of IPTG and the cultures were incubated at 25 ºC with 175 rpm shaking until OD_{600} reached 0.7. Each culture was sparged with argon for 20 min and brought into the anaerobic chamber kept at 4 ºC (Coy Laboratory Products). The cells were centrifuged under anaerobic condition at 3,200 ×*g* for 15 min at 4 ºC and the supernatant was decanted. The pelleted cells were resuspended in 1 mL of anoxic lysis buffer (20 mM

HEPES at pH 8, 100 mM NaCl, 10 $MgCl_2$, 10 mM DTT) supplemented with EDTA-free Pierce Protease Inhibitor Tablet (Thermo Fisher). The resuspended cells were lysed by sonication on ice (Branson Digital Sonifier, sonicator setting: 25% amplitude, total time = 1 min, on cycle = 10 sec, off cycle = 30 sec). The resulting lysate was used directly for *in vitro* assays.

The assay mixture containing 200 µM substrate (resorcinol **4.28** or cylindrocyclophane F **4.7** as DMSO stock), 100 µg/mL spinach ferredoxin, 100 mU/mL spinach ferredoxin reductase, 1 mM NADPH and lysate in assay volume of 100 µL was prepared. The assay mixtures were incubated at room temperature for 18 h under aerobic condition. The assay samples were quenched with 100 µL of cold acetonitrile and 100 µL of methanol. The quenched samples were centrifuged at 16,100 □*g* for 10 min and the supernatant was analyzed by LC-HRMS and LC-HRMS/MS.

5 µL samples were analyzed by LC-HRMS on a Gemini C18 analytical column (4.6 x 50 mm, 5 µm) equipped with a guard column at a flow rate of 0.4 mL/min with the column temperature maintained at room temperature. The following elution conditions were applied: 75% solvent B in solvent A for 2 min, a gradient increasing from 75% to 100% solvent B in solvent A over 15 min, 100% solvent B for 3 min, a gradient decreasing to 100% to 75% solvent B in solvent A over 0.1 min, and 75% solvent B in solvent A for 10 min. (solvent A = 95:5 water/acetonitrile with 10 mM ammonium hydroxide; solvent B = 92:2 acetonitrile/water in 10 mM ammonium hydroxide). For the MS detection, the ESI mass spectra data were recorded on a negative ionization mode for a mass range of m/z 50 to 3000. A mass window of ± 0.005 Da was used to extract the ion of $[M-H]^-$. For ions of interest, MS/MS analysis was performed using fragmentation voltage of 30.0 eV.

4.4.5. HHpred search and generation of homology model for CylC

CylC was used as the query to search for structural homologs using HHpred.[22] The protein from the PDB database with the best homology was AurF[23] and a homology model of CylC was generated by Modeller (Max-Plank Institute for Developmental Biology)[29] using the crystal structure of AurF as a template and

the resulting PDB file was aligned with the AurF PDB file (3CHH) using MacPyMOL version 1.7 (Schrödinger Inc). The residues required for binding the metal co-factor in AurF was visualized, and the corresponding residues in the homology model of CylC were identified.

4.4.6. Multiple sequence alignment of CylC and its homologs

BLASTp search was conducted with CylC as a query to return 33 close homologs as hits. The sequences of CylC homologs, including BrtJ, ColD, ColE and CylC homolog in the microginin biosynthetic gene cluster, were used to generate a multiple sequence alignment using ClustalW on Geneious Pro Version 7.1.6 (Biomatters). The multiple sequence alignment of 18 out of 33 sequences, including CylC, is shown in **Figure 4.3C**.

4.4.7. Cloning of *brtJ*

Table 4.4. Oligonucleotides used for cloning *brtJ* (restriction sites underlined).

Primer name	Sequence (5' to 3')	Target
BrtJ-F-BsaI-pBR-IBA2	ATGATCGGTCTCTGCGCCGTTTCTGTAACGCCATC	N-Strep-BrtJ
BrtJ-R-BsaI-pBR-IBA2	AATCACGGTCTCATATCTTACACGGTTCTAGCG	

BrtJ was PCR amplified from *Synechocystis salina* LEGE 06155 genomic DNA using the primers shown in **Table 4.4**. Both forward and reverse primers have a *Bsa*I restriction site. The PCR mixtures contained Phusion High-Fidelity PCR Master Mix, 70 ng DNA template, and 25 pmoles of each primer in a total volume of 50 μL. The thermocycling was carried out using the following parameters: denaturation for 30 sec at 98 ºC, followed by 35 cycles of 10 sec at 98 ºC, 30 sec at 60 ºC, 1.5 min at 72 ºC, and a final extension time of 10 min at 72 ºC. PCR products were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled, and purified.

Amplified fragments were digested with *Bsa*I for 2.5 h at 37 ºC. Digests contained 4 µL water, 3 µL Cutsmart buffer (10x), 20 µL of PCR product, 3 µL of *Bsa*I (10 U/µL). Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digests were ligated into linearized pPR-IBA2 expression vector using T4 DNA ligase. Ligation samples were incubated at 16 ºC overnight and contained 3 µL water, 1 µL T4 Ligase Buffer (10x), 1 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA Ligase (400 U/µL). The ligation sample (5 µL) was used to transform a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pPR-IBA2-*brtJ* construct was confirmed by sequencing of the purified plasmid DNA.

4.4.8. Expression and purification of BrtJ

The pPR-IBA2-*brtJ* was transformed into chemically competent *E. coli* Rosetta (DE3) cells (Invitrogen), and the resulting colonies were used directly to prepare 50 mL starter culture in LB medium supplemented with 50 µg/mL ampicillin, 34 µg/mL chloramphenicol and 1g/L glucose. The starter culture was incubated overnight at 37 ºC. Overnight cultures were diluted 1:100 into 2 L LB medium containing 50 µg/mL ampicillin, 34 µg/mL chloramphenicol and 1g/L glucose. The cultures were incubated at 37 ºC with 175 rpm shaking. At OD_{600} = 0.1, 100 µM manganese chloride and 100 µM ammonium iron citrate were added, and the cultures were incubated further at 37 ºC with 175 rpm shaking. When the cultures reached OD_{600} = 0.35–0.4, the cultures were cooled to 25 ºC, and at OD_{600} = 0.5–0.6, the protein expression was induced with 100 µM IPTG. The cultures were incubated for 19 h at 25 ºC with 175 rpm shaking.

Cells from 2 L of culture were pelleted by centrifugation (6,720 □*g* for 15 min) and resuspended in 50 mL lysis buffer (20 mM EPPS, 500 mM NaCl, 10 mM $MgCl_2$, pH 8.7 or 100 mM potassium phosphate, 500 mM NaCl, pH 8 with 1% Triton X-100 or 8 mM CHAPS) supplemented with EDTA-free Pierce Protease Inhibitor Tablets (Thermo Fisher). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 8,000–10,000 psi, and the lysate was clarified by centrifugation (28,900 □*g* for

30 min). The supernatant was incubated with 2 mL of *Strep*-Tactin resin for 2 h at 4 ºC. The mixture was centrifuged (3,200 □*g* for 5 min) and the unbound fraction was discarded. The *Strep*-Tactin resin was resuspended in 5 mL of wash buffer (lysis buffer lacking detergent), loaded into a glass column, and washed with 12 mL of wash buffer. Protein was eluted from the column in elution buffer (lysis buffer containing 2.5 mM desthiobiotin) in 1 mL fractions. SDS-PAGE analysis (4–15% Mini-PROTEAN TGX precast gel) was employed to determine the presence and purity of protein in each fraction. Fractions containing the desired protein were combined and dialyzed twice against 1 L of dialysis buffer (25 mM EPPS, 100 mM NaCl, 10% glycerol, pH 8.7 or 50 mM potassium phosphate, 300 mM NaCl, 10% glycerol, pH 8.0). The dialyzed protein solution was concentrated using a Spin-X UF 20 mL centrifugal concentrator with a 10,000 MWCO membrane (Corning), and the concentrated protein solutions were frozen with liquid nitrogen for storage at –80 ºC. This procedure afforded 1.5 mg/L of BrtJ using EPPS lysis buffer and 0.5 mg/L of BrtJ using Triton X-100 or 1.5 mg/L of BrtJ using CHAPS in the potassium phosphate lysis buffer.

4.4.9. Cloning of CylA, CylB and CylC for co-expression

Table 4.5. Oligonucleotides used for cloning *cylA, cylB* and *cylC* (restriction sites underlined).

Primer name	Sequence (5' to 3')	Target
Duet-CylA-NdeI-F	CACCGGCGACATATGCATTTGCTACAAC	CylA
Duet-CylA-XhoI-R	CGGCGGCTCGAGTCATTCTCTTTTTTTGG	
Duet-CylB-BamHI-F	CGAGTGGGATCCGGAAATCGTAAATAG	N-His_6-CylB
Duet-CylB-HindIII-R	CGGCGGAAGCTTTCATGAATATTCTGC	
Duet-CylC-MCS1-BamHI-F	GAGGTCGGATCCCATGACTAGCGTTAATG	N-His_6-CylC
Duet-CylC-MCS1-PstI-R	GACAGTCTGCAGCTATTTTATAGCCCCC	
Duet-CylC-MCS2-BglII-F	GGCGGCAGATCTCATGACTAGCGTTAATG	CylC
Duet-CylC-MCS2-XhoI-R	GTCCATCTCGAGCTATTTTATAGCCCCC	

CylA, cylB and *cylC* were PCR amplified from the *Cylindrospermum licheniforme* ATCC 29412 genomic DNA using the primers shown in **Table 4.5**. The restriction enzymes used for each primer is indicated in the primer names. The PCR mixture contained Phusion High-Fidelity PCR Master Mix, 50 ng DNA template, and 25 pmoles of each primer in a total volume of 50 µL. The thermocycling was carried out using the following parameters: denaturation for 30 sec at 98 ºC, followed by 35 cycles of 10 sec at 98 ºC, 30 sec at 44 ºC, 1.75 min at 72 ºC, and a final extension time of 10 min at 72 ºC. PCR products were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled, and purified.

Amplified fragments were digested with restriction enzymes (*cylA* = *Nde*I and *Xho*I, N-His$_6$-*cylB* = *Bam*HI-HF and *Hind*III-HF, NHis$_6$-CylC = *Bam*HI and *Pst*I, CylC = *Bgl*II and *Xho*I) for 2.5 h at 37 ºC. Digests contained 4 µL water, 3 µL NEB buffer (10x Cutsmart buffer for *cylA* and *cylB,* 10x NEB buffer 3.1 for *cylC*), 20 µL of PCR product, 1.5 µL each of restriction enzymes (20 U/µL). Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digests of NHis$_6$-*cylC* or *cylC* were ligated into the multiple cloning site (MSC) 1 or 2 of the linearized pETDuet expression vector, respectively. The digest of *cylA* was ligated into MCS2 of the linearized pCOLADuet vector. Ligations were incubated at 16 ºC overnight and contained 3 µL water, 1 µL T4 Ligase Buffer (10x), 1 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA Ligase (400 U/µL). The ligation sample (5 µL) was used to transform a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pETDuet-NHis$_6$-*cylC*, pETDuet-*cylC* and pCOLADuet-*cylA* constructs were confirmed by sequencing of the purified plasmid DNA.

Next, the pCOLADuet-*cylA* vector was linearized with *Bam*HI-HF and *Hind*III-HF for 2.5 h at 37 ºC. Digests contained 7 µL water, 3 µL Cutsmart buffer (10x), 16 µL of plasmid DNA, and 1.5 µL each of restriction enzymes (20 U/µL). Then, 1 µL of calf intestinal alkaline phosphatase (CIP, 10 U/µL) was added and further incubated at 37 ºC for 30 min. Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digested NHis$_6$-*cylB* was ligated into the MCS1 of the linearized pCOLADuet-*cylA* expression vector. Ligations were

incubated at 16 ºC overnight and contained 3 µL water, 1 µL T4 Ligase Buffer (10x), 1 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA Ligase (400 U/µL). The ligation sample (5 µL) was transferred to a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pCOLADuet-N-His_6-*cylB-cylA* construct was confirmed by sequencing of the purified plasmid DNA.

4.4.10. Expression and purification of CylB and CylC from co-expression strain

pETDuet-N-His_6-*cylC* or pETDuet-*cylC* and pCOLADuet-N-His_6-*cylB-cylA* were co-transformed into electrocompetent BAP1 *E. coli* cells.[40] The freshly transformed colonies (less than 1 day old) were picked to inoculate 25 mL starter cultures in LB medium supplemented with 50 µg/mL ampicillin and 30 µg/mL kanamycin. The starter culture was incubated overnight at 37 ºC. Overnight cultures were diluted 1:100 into 1 L LB medium containing ampicillin and kanamycin. The cultures were incubated at 37 ºC with 190 rpm shaking. At OD_{600} = 0.3–0.4, the cultures were cooled to 25 ºC, and at OD_{600} = 0.5–0.7, the protein expression was induced with 100 µM IPTG and the cultures were supplemented with 1 mM decanoic acid (**4.1**, 1 M stock in DMSO). The cultures were incubated for 20 h at 25 ºC with 175 rpm shaking.

Cells from 1 L of culture were pelleted by centrifugation (6,720 □*g* for 15 min) and resuspended in 25 mL lysis buffer (20 mM HEPES or Tris, 500 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP) supplemented with EDTA-free Pierce Protease Inhibitor Tablets (Thermo Fisher). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 8,000–10,000 psi, and the lysate was clarified by centrifugation (28,900 □*g* for 30 min). The supernatant was incubated with 2 mL of Ni-NTA resin with 10 mM imidazole for 2 h at 4 ºC. The mixture was centrifuged (3,200 □*g* for 5 min) and the unbound fraction was discarded. The Ni-NTA resin was resuspended in 5 mL of wash buffer 1 (lysis buffer containing 20 mM imidazole), loaded into a glass column, and washed with 8 mL of wash buffer 1 and 6 mL of wash buffer 2 (lysis buffer containing 40 mM imidazole). Bound protein was eluted from the column in elution buffer (lysis buffer containing 200 mM imidazole) collecting 2 mL fractions. SDS-PAGE analysis (4–

15% Mini-PROTEAN TGX precast gel) was employed to determine the presence and purity of protein in each fraction. Fractions containing the desired protein were combined and dialyzed twice against 2 L of dialysis buffer (20 mM HEPES or Tris, 50 mM NaCl, 1 mM TCEP, 10% glycerol, pH 8.0). The dialyzed protein solution was concentrated using a Spin-X UF 20 mL centrifugal concentrator with a 5,000 MWCO membrane (Corning), and the concentrated protein solutions were frozen with liquid nitrogen for storage at –80 ºC. This procedure afforded 8.9 mg/L or 11.4 mg/L of protein mixture containing mostly $NHis_6$-CylB and untagged CylC or $NHis_6$-CylC, respectively.

4.4.11. Gel filtration FPLC analysis of BrtJ and CylC

A 50 μM solution of N-Strep-BrtJ and 5 mg/mL solution of N-His_6-CylB and CylC (untagged or $NHis_6$ tagged) was analyzed by gel filtration. N-Strep-BrtJ was eluted in EPPS buffer (25 mM EPPS, pH 8.7, 100 mM NaCl, 10% glycerol) or phosphate buffer (50 mM potassium phosphate buffer, pH 8, 100 mM NaCl containing 10% glycerol), and N-His_6-CylB/CylC mixture was eluted in 20 mM Tris buffer, pH 8, 50 mM NaCl containing 10% glycerol at 0.25 mL/min using Superdex 200 30/100 GL column. A solution of molecular weight markers was analyzed under the same conditions.

4.4.12. Ferene S spectrophotometric assay of BrtJ and CylC

The iron content of a 49.2 μM N-Strep-BrtJ in EPPS buffer at pH 8.7, a protein mixture containing 5.9 mg/ml of untagged CylC and N-His_6-CylB, and a protein mixture containing 3.6 mg/mL of N-His_6-CylC and N-His_6-CylB was determined using Ferene (3-(2-pyridyl)-5,6-di(2-furyl)-1,2,4-triazine-5, 5”-disulfonic acid disodium salt), according to a previously published procedure,[42] with the exceptions that the assay volume was reduced to 120 μL total and the standard curve was prepared with ammonium iron (II) sulfate hexahydrate (0–50 or 0–100 μM). The absorbance at 592 nm was measured at 25 ºC in 96-well plates using a PowerWave HT Microplate Spectrophotometer (BioTek). Path length correction was employed such that absorbance values were reported for a path length of 1 cm. Data was analyzed using Gen5 Data Analysis Software.

4.4.13. LC-HRMS analysis of the trypsin digested CylB from the co-expression strain

A solution of 1 mg/mL N-His_6-CylB + N-His_6-CylC or 1 mg/mL N-His_6-CylB purified from BAP1 *E. coli* strain lacking *cylC* expression vector was prepared in buffer containing 50 mM Tris, 100 mM NaCl and 2 mM $MgCl_2$ at pH 8. The standard of 6-chlorodecanoyl-CylB (**4.67**) was prepared by incubating 100 µM *apo* N-His_6-CylB (purified following the procedure in **Chapter 2.4.6**) with 5 µM Sfp, 200 µM 6-chlorodecanoyl-CoA (**4.85**) and 2 mM $MgCl_2$ in 50 mM Tris-HCl buffer (pH 8.0) for 1 h at room temperature prior to trypsin digest.

Trypsin digest of the assay mixture and the 6-chlorodecanoyl-CylB standard was performed following the procedure described in the manual provided by Promega. Aliquots (25 µL) of 0.1 µg/µL Sequencing Grade Modified Trypsin (Promega) were incubated at 30 ºC for 15 min. Buffer A (12.5 µL, 0.1 M ammonium bicarbonate, 2 mM TCEP, pH 7), 12.5 µL of protein sample and 25 µL of pre-warmed trypsin were mixed and incubated at 30 ºC for 2.5 h. The reaction was quenched with 25 µL of 25% v/v formic acid, and the samples were diluted with 75 µL of water. The reaction was centrifuged at 16,100 □*g* for 10 min and the supernatant was submitted for LC-HRMS analysis.

The theoretical mass of the trypsin-digested peptide containing the active site serine residue that carries the ppant arm was calculated using the ProteinProspector online tool. The analysis was performed on Thermo q-Exactive Plus equipped with Dionex Ultimate 3000 HPLC (Thermo Scientific) in the Small Molecule Mass Spectrometry Facility, FAS Division of Science.

4.4.14. CylA fatty acid activation assays using pre-functionalized substrates

A stock solution of *holo*-CylB was prepared prior to the assay by incubating 200 µM *apo*-CylB, 8 µM Sfp, 6 mM CoA, and 2 mM $MgCl_2$ in 50 mM Tris-HCl pH 8.0 for 1 h at room temperature. Loading assays contained 50 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 100 µM ATP, 50 µM decanoic acid (**4.1**, **4.73** or **4.74**) added as a 5 mM stock solution in DMSO, 50 µM *holo*-CylB and 100 nM CylA in a final assay volume of 100 µL. After 2.5 h, samples were quenched with equal volume of acetonitrile and centrifuged

(16,100 ×*g* for 10 min at 4 ºC) before HPLC analysis on a Kinetex C8 column (2.6 µm, 100 Å, 100 × 3.0 mm, Phenomenex) at a flow rate of 0.5 mL/min. HPLC assay conditions were: a gradient increasing from 0–100% B in solvent A over 30 min, 100% B for 5 min, a gradient decreasing to 0% B in solvent A over 2 min, 0% B for 8 min (solvent A = 0.1% TFA in water, solvent B = 0.1% TFA in acetonitrile).

4.4.15. CylI assay with non-chlorinated and chlorinated β-ketoacyl-SNAC substrate

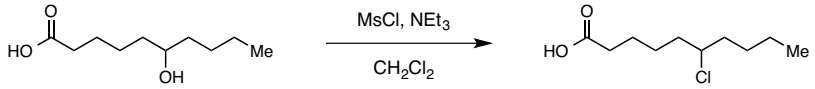
Each assay mixture contained 100 mM potassium phosphate buffer pH 8.0, 1 mM EDTA, 250 µM malonyl-CoA, 100 µM substrate (**4.80** or **4.81**) and 1.2 µM N-His_6-CylI in a final assay volume of 100 µL. An aqueous solution containing potassium phosphate buffer, EDTA, substrate and malonyl-CoA was prepared first. The mixture was vortexed, the enzyme was added, and the reaction was mixed gently with a pipet and incubated at room temperature. After 5 min, 15 min and 1 h, 30 µL aliquots were removed from the reaction mixture, added to 60 µL of ice-cold acetonitrile, vortexed, incubated on ice for 10 min and centrifuged (16,100 ×*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC (80 µL injection volume) on a Chromolith RP-18 endcapped monolithic silica column (4.6 x 100 mm, EMD Millipore) at a flow rate of 1 mL/min. HPLC assay conditions were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 15 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 6 min (solvent A = water, solvent B = acetonitrile).

4.4.16. Coupled $CylH_{PKS}$ and CylI assay with non-chlorinated or chlorinated acyl-SNAC substrate

Stock solution of *holo* $CylH_{PKS}$ was prepared prior to the assay by incubating 50 µM *apo* $CylH_{PKS,}$ 5 µM Sfp, 250 µM CoA and 1 mM $MgCl_2$ in 50 mM potassium phosphate buffer at pH 8 in 50 µL volume for 1 h at room temperature. Each CylI assay contained 100 µM substrate (**4.76** or **4.77**), 20 µM *holo* $CylH_{PKS}$, 1.2 µM CylI, 500 µM malonyl-CoA, and 1 mM EDTA in 100 mM potassium phosphate buffer at pH 8 in a final assay volume of 75 µL. The reaction was incubated at room temperature and 30 µL of the reaction was quenched with 60 µL of cold acetonitrile after 2 and 5 h. The quenched sample was incubated on ice for 10 min and then centrifuged (16,100 ×*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC

(80 µL injection volume) on a Chromolith RP-18 endcapped monolithic silica column (4.6 x 100 mm, EMD Millipore) with or without a guard column at a flow rate of 1 mL/min. HPLC assay conditions were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 15 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 6 min (solvent A = water, solvent B = acetonitrile).

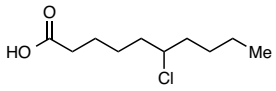
4.4.17. Synthesis of the functionalized decanoic acid derivatives for CylA assays



4.73 *78%* **4.74**

Scheme 4.14: Synthesis of 6-chlorodecanoic acid.

The synthesis of 6-hydroxydecanoic acid (**4.73**) was performed following previously reported procedures.[59,60]

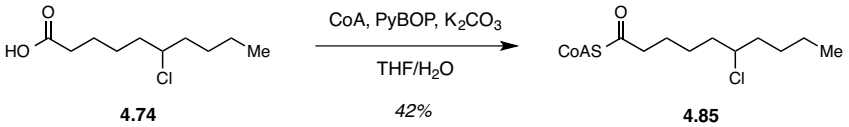


4.74

6-Chlorodecanoic acid (4.74): 6-Hydroxydecanoic acid **4.73** (415 mg, 2.2 mmol) was dissolved in dichloromethane (10 mL) and cooled to 0 ºC. Mesyl chloride (378 µL, 4.9 mmol) and triethylamine (1.54 mL, 11.0 mmol) were added, and the reaction mixture was stirred at 0 ºC for 30 min. The reaction was quenched with 1 M KOH (10 mL) and extracted with dichloromethane (3 × 10 mL). The combined organic layers were washed with brine (50 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The resulting mesylate was dissolved in THF (10 mL), and anhydrous lithium chloride (466 mg, 11.0 mmol) was added. The reaction mixture was heated to reflux overnight. The reaction was cooled to room temperature, diluted with water (20 mL), acidified with 1 M HCl (5 mL), and extracted with ethyl acetate (3 × 20 mL). The combined organic layers were washed with brine (50 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 20-40% ethyl acetate in hexanes to afford 6-chlorodecanoic acid **4.74** (353 mg, 1.7 mmol, 78%). TLC: R_f = 0.30 (4:1 hexanes/ethyl acetate). ^{1}H-NMR (500 MHz; $CDCl_3$):

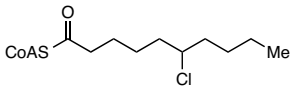
δ 11.49 (s, 1H, COO**H**), 3.88 (m, 1H, C**H**Cl), 2.37 (t, 2H, *J* = 7.3 Hz, COC**H**$_2$), 1.78-1.53 (m, 8H, C**H**$_2$), 1.53-1.41 (m, 2H, C**H**$_2$), 1.41-1.26 (m, 2H, C**H**$_2$), 0.90 (t, 3H, *J* = 7.2 Hz, C**H**$_3$). ^{13}C NMR (125 MHz; $CDCl_3$): δ 180.3, 64.0, 38.5, 38.3, 34.2, 28.9, 26.2, 24.4, 22.5, 14.2.

4.4.18. Synthesis of 6-chlorodecanoyl-CoA for the preparation of 6-chlorodecanoyl-CylB standard



4.74 **4.85**

Scheme 4.15: Synthesis of 6-chlorodecanoyl-CoA.

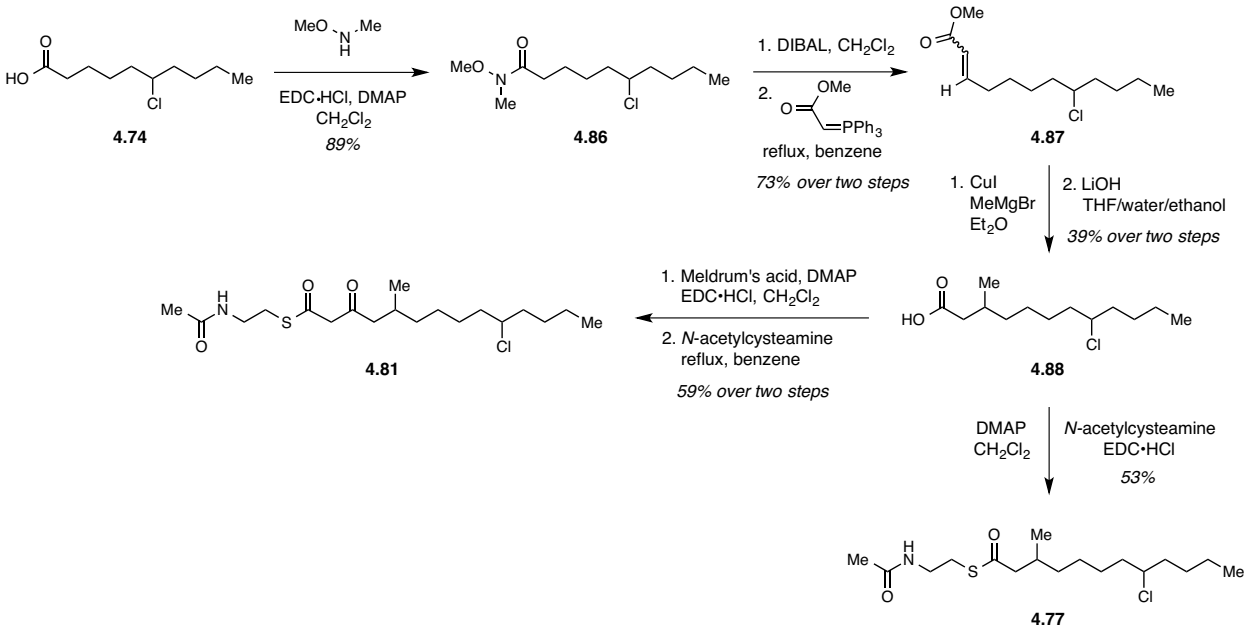


4.85

6-Chlorodecanoyl-CoA (4.85): 6-Chlorodecanoic acid **4.74** (5.2 mg, 0.025 mmol), CoA hydrate (21 mg, 0.028 mmol), PyBOP (26 mg, 0.05 mmol), and potassium carbonate (14 mg, 0.1 mmol) were dissolved in a 1:1 solution of THF/water (1 mL). The reaction was stirred at room temperature for 2 h. The crude product was centrifuged at 16,100 □*g* for 10 min and the supernatant was purified by preparative HPLC. The purification was performed using a Kromasil 100 C18 column (5 μm, 10 x 150 mm) at a flow rate of 3 mL/min. HPLC conditions were: a gradient increasing from 5% to 95% solvent B in solvent A over 30 min, 95% solvent B for 10 min, a gradient decreasing to 5% solvent B over 5 min, 5% solvent B for 10 min (solvent A = water with 0.1% formic acid, solvent B = acetonitrile with 0.1% formic acid). 6-Chlorodecanoyl-CoA **4.85** eluted at 15-20 min, and the collected fractions were lyophilized overnight (10 mg, 0.010 mmol, 42%). HRMS (ESI): calc'd for $C_{31}H_{52}ClN_7O_{17}P_3S^-$ $[M–H]^-$, 954.2047; found, 954.2047. ^{1}H-NMR (500 MHz; CD_3OD): 8.69 (s, 1H, C**H**), 8.36 (s, 1H, C**H**), 6.11 (d, 1H, *J* = 5.2 Hz, NOC**H**), 4.98 (d, 1H, *J* = 8.6, 4.1 Hz, C**H**OH), 4.81 (m, 1H, C**H**OPO$_3^{2-}$), 4.54-4.51 (m, 1H, OC**H**), 4.40 (d, 1H, *J* = 12.0 Hz, O$_3^{2-}$POCH**H**), 4.33 (d, 1H, *J* = 12.0 Hz, O$_3^{2-}$POCH**H**), 4.09 (dd, 1H, *J* = 9.6, 4.9 Hz, OC**H**$_2$C(CH$_3$)$_2$), 4.03 (s, 1H, C**H**OH), 3.90 (m, 1H, C**H**Cl), 3.71 (dd, 1H, *J* = 9.5, 4.3 Hz, OC**H**$_2$C(CH$_3$)$_2$), 3.50-3.41 (m, 4H, NHC**H**$_2$), 2.98 (t, 2H, *J* = 6.8 Hz, SCOC**H**$_2$), 2.59 (t, 2H, *J* = 7.3 Hz, NHCOC**H**$_2$), 2.41

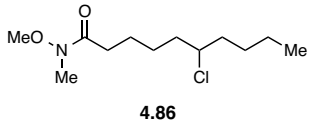
(t, 2H, J = 6.8 Hz, SC**H**$_2$), 1.79-1.70 (m, 2H, C**H**$_2$), 1.70-1.59 (m, 4H, C**H**$_2$), 1.59-1.26 (m, 6H, C**H**$_2$), 1.07 (s, 3H, C(C**H**$_3$)$_2$), 0.92 (t, 3H, J = 7.2 Hz, CH$_2$C**H**$_3$), 0.88 (s, 3H, C(C**H**$_3$)$_2$).

4.4.19. Synthesis of the chlorinated substrate for CylI and CylH$_{PKS}$ assays



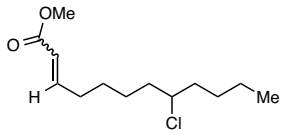
Scheme 4.16: Synthesis of chlorinated SNAC substrates for CylI and CylH$_{PKS}$.

6-chlorodecanoic acid (**4.74**) was synthesized following the procedure described in section **2.4.16**.



Weinreb amide (4.86): 6-Chlorodecanoic acid **4.74** (550 mg, 2.66 mmol) was dissolved in anhydrous dichloromethane (25 mL) under argon. EDC•HCl (765 mg, 3.99 mmol), DMAP (488 mg, 3.99 mmol) and *N*,*O*-dimethylhydroxylamine hydrochloride (389 mg, 3.99 mmol) were added and the reaction mixture was stirred at room temperature overnight. The reaction mixture was quenched with brine (30 mL), and the organic layer was separated. The aqueous layer was extracted with ethyl acetate (3 × 30 mL). The combined organic layers were washed with 2 M HCl (30 mL) and brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to obtain Weinreb amide **4.86** (589 mg, 2.36 mmol,

89%). The product was used for the next reaction without further purification. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.90 (m, 1H, C**H**Cl), 3.69 (s, 3H, OC**H**$_3$), 3.18 (s, 3H, NC**H**$_3$), 2.44 (t, 2H, J = 7.1 Hz, COC**H**$_2$), 1.80-1.55 (m, 6H, C**H**$_2$), 1.54-1.44 (m, 2H, C**H**$_2$), 1.43-1.25 (m, 4H, C**H**$_2$), 0.91 (t, 3H, J = 7.2 Hz, C**H**$_3$).

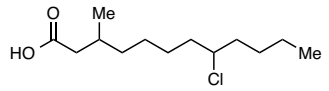


4.87

Methyl 8-chlorododec-2-enoate (4.87): Weinreb amide **4.86** (589 mg, 2.36 mmol) was dissolved in anhydrous dichloromethane (8 mL), and the solution was cooled to –78 ºC. To the stirring solution, DIBAL (1 M solution in dichloromethane, 4.72 ml, 4.72 mmol) was added dropwise over 5 min and the reaction mixture was stirred at the same temperature for 1 h. The reaction mixture was quenched by addition of the saturated aqueous solution of potassium sodium tartrate (10 mL). The mixture was slowly warmed to room temperature and stirred for 30 min before diluting with brine (20 mL). The aqueous layer was extracted with ethyl acetate (3 × 25 mL), and the combined organic layers were washed with brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was further purified by flash chromatography, eluting with 5-15% ethyl acetate in hexanes to afford 6-chlorodecanal as a volatile colorless oil. The product was carried on to the next reaction without further purification. TLC: R_f = 0.83 (silica gel, 1:1 hexanes/ethyl acetate). ^{1}H-NMR (600 MHz; $CDCl_3$): δ 9.72 (s, 1H, OC**H**), 3.84 (m, 1H, C**H**Cl), 2.42 (td, 2H, J = 7.2, 1.6 Hz, COC**H**$_2$), 1.70-1.26 (m, 12H, C**H**$_2$), 0.87 (t, 3H, J = 7.3 Hz, C**H**$_3$).

Crude 6-chlorodecanal (2.36 mmol) was dissolved in anhydrous benzene (10 mL). Methyl (triphenylphosphoranylidene)acetate (0.946 g, 2.83 mmol) was added to the reaction flask and the reaction mixture was refluxed for 4 h. The reaction was concentrated *in vacuo*, and the crude product was purified by flash chromatography, eluting with 0-10% ethyl acetate in hexanes to yield methyl 8-chlorododec-2-enoate **4.87** as a colorless oil and a mixture of isomers in a *Z*:*E* = 1:20 ratio (423 mg, 1.71

mmol, 73%). TLC: R_f = 0.48 (*Z*), 0.43 (*E*) (silica gel, 9:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{13}H_{24}ClO_2^+$ $[M+H]^+$, 247.1459; found, 247.1474. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 6.95 (dt, 1H, *J* = 15.6, 7.0 Hz, C**H**CH_2), 5.82 (dt, 1H, *J* = 15.6, 1.6 Hz, COC**H**), 3.86 (m, 1H, C**H**Cl), 3.71 (s, 3H, OC**H**$_3$), 2.23-2.19 (m, 2H, CHC**H**$_2$), 1.72-1.67 (m, 4H, CHClC**H**$_2$), 1.51-1.29 (m, 8H, C**H**$_2$), 0.90 (t, 3H, *J* = 7.2 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 167.0, 149.1, 121.1, 63.9, 51.3, 38.2(2), 32.0, 28.6, 27.6, 26.0, 22.2, 13.9.

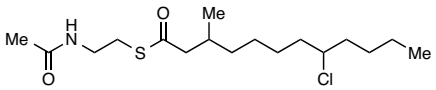


4.88

8-Chloro-3-methyldodecanoic acid (4.88): To a suspension of copper iodide (200 mg, 1.05 mmol) in anhydrous diethyl ether (5 mL) at –20 ºC, methylmagnesium bromide (3 M in diethyl ether, 1.73 ml, 5.19 mmol) was added dropwise over 15 min, and the reaction mixture was stirred at the same temperature for 1.5 h. Then, a solution of methyl 8-chlorododec-2-enoate **4.87** (256 mg, 1.04 mmol) in anhydrous ether (3 mL) was added dropwise over 5 min, and the reaction mixture was stirred at –20 ºC for an additional 1.5 h. The reaction mixture was quenched by slow addition of methanol (1 mL) followed by an aqueous solution of saturated sodium bicarbonate (40 mL). The aqueous layer was extracted with ethyl acetate (3 × 50 mL), and the combined organic layers were washed with brine (40 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude methyl 8-chloro3-methyldodecanoate was used directly in the next reaction without further purification. TLC: R_f = 0.71 (silica gel, 9:1 hexanes/ethyl acetate).

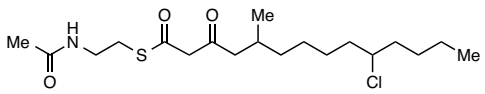
The crude mixture of methyl 8-chloro3-methyldodecanoate (1.05 mmol) was dissolved in 1:1:1 solution of THF/water/ethanol (15 mL), and lithium hydroxide (30.0 mg, 1.25 mmol) was added. The reaction mixture was stirred at room temperature for 5 h. The reaction mixture was concentrated *in vacuo*, acidified with 1 M HCl (20 mL), and the aqueous layer was extracted with dichloromethane (3 × 25 mL). The combined organic layers were washed with brine (40 mL), dried over anhydrous sodium sulfate,

filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography and the desired product was eluted with 20-50% ethyl acetate in hexanes to afford 8-chloro-3-methyldodecanoic acid **4.88** as a colorless oil (117 mg, 0.47 mmol, 39%). TLC: R_f = 0.5 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{13}H_{24}ClO_2^-$ $[M–H]^-$, 247.1470; found 247.1470. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 12.0-11.0 (bs, 1H, COO**H**), 3.88 (m, 1H, C**H**Cl), 2.34 (dd, 1H, J = 15.0, 6.0 Hz, COC**H**$_2$), 2.15 (dd, 1H, J = 15.0, 8.1 Hz, COC**H**$_2$), 1.97-1.94 (m, 1H, C**H**CH$_3$), 1.74-1.64 (m, 4H, CHClC**H**$_2$), 1.54-1.18 (m, 10H, C**H**$_2$), 0.96 (d, 3H, J = 6.7 Hz, CHC**H**$_3$), 0.90 (t, 3H, J = 7.2 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 179.8, 64.2, 41.5, 38.4, 38.2, 36.4, 30.1, 28.7, 26.6, 26.5, 22.3, 19.7, 14.0.



4.77

8-Chloro-3-methyldodecanoyl-SNAC thioester (4.77): 8-Chloro-3-methyldodecanoic acid **4.88** (33 mg, 0.13 mmol) was dissolved in dichloromethane (2.5 mL), and the solution was cooled to 0 ºC. EDC•HCl (51 mg, 0.27 mmol) was added to the reaction flask, and the reaction mixture was stirred at 0 ºC for 20 min. Then, *N*-acetylcysteamine (16 μL, 0.15 mmol) and DMAP (8 mg, 0.07 mmol) were added, and the reaction mixture was warmed to room temperature and stirred for 3 h. The reaction mixture was quenched with water (10 mL) and extracted with ethyl acetate (3 × 10 mL). The combined organic layers were washed with brine (20 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography eluting at 40-50% ethyl acetate in hexanes to afford 8-chloro-3-methyldodecanoyl-SNAC thioester **4.77** as a white solid (25 mg, 53%). TLC: R_f = 0.24 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{17}H_{33}ClNO_2S^+$ $[M+H]^+$, 350.1915; found 350.1931. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.90 (bs, 1H, N**H**), 3.87 (m, 1H, C**H**Cl), 3.42 (q, 2H, J = 6.2 Hz, NHC**H**$_2$), 3.02 (t, 2H, J = 6.4 Hz, CH$_2$C**H**$_2$S), 2.55 (dd, 1H, J = 14.6, 6.0 Hz, SCOC**H**$_2$), 2.38 (dd, 1H, J = 14.6, 8.1 Hz, SCOC**H**$_2$), 2.01 (m, 1H, C**H**), 1.97 (s, 3H, NHCOC**H**$_3$), 1.74-1.64 (m, 4H, C**H**$_2$CHCl), 1.53-1.44 (m, 2H, C**H**$_2$), 1.42-1.17 (m, 8H, C**H**$_2$), 0.90 (m, 6H, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 199.7, 170.2, 64.2, 51.3, 39.8, 38.4, 38.2, 36.4, 31.0, 28.6, 28.5, 26.5, 26.4, 23.2, 22.3, 19.5, 14.0.



4.81

10-Chloro-5-methyl-3-oxotetradecanoyl-SNAC thioester (4.81): 8-Chloro-3-methyldodecanoic acid **4.88** (117 mg, 0.47 mmol) was dissolved in anhydrous dichloromethane (5 mL), and to this solution, Meldrum's acid (67.8 mg, 0.47 mmol) and DMAP (172 mg, 1.41 mmol) were added. The reaction mixture was cooled to 0 ºC, and a solution of EDC•HCl (178 mg, 0.94 mmol) in dichloromethane (1 mL) was added dropwise over 5 min. The reaction mixture was warmed to room temperature and stirred overnight. The reaction was diluted with dichloromethane (20 mL), and the organic layer was washed with 1 M HCl (3 × 30 mL) and brine (30 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to obtain crude 5-(8-chloro-3-methyldodecanoyl)-2,2-dimethyl-1,3-dioxane-4,6-dione in quantitative yield. The crude product was used directly in the next reaction without further purification. TLC: R_f = 0.25 (silica gel, 1:1 hexanes/ethyl acetate).

5-(8-Chloro-3-methyldodecanoyl)-2,2-dimethyl-1,3-dioxane-4,6-dione (0.47 mmol) was dissolved in benzene (5 mL) and *N*-acetylcysteamine (101 µL, 0.93 mmol) was added. The reaction mixture was refluxed overnight. The reaction mixture was then cooled to room temperature, concentrated *in vacuo*, and purified by flash chromatography, eluting with 50-70% ethyl acetate in hexanes to afford 10-chloro-5-methyl-3-oxotetradecanoyl-SNAC thioester **4.81** as a colorless oil and a mixture of tautomers (110 mg, 0.28 mmol, 59%). TLC: R_f = 0.15 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{19}H_{33}ClNO_3S^-$ $[M–H]^-$, 390.1875; found 390.1878. ^{1}H-NMR (500 MHz; $CDCl_3$): **Keto form:** δ 5.92 (s, 1H, N**H**), 3.89 (m, 1H, C**H**Cl), 3.50-3.43 (m, 2H, NHC**H**$_2$), 3.12-3.07 (m, 2H, SC**H**$_2$), 2.51 (dd, 1H, *J* = 16.6, 5.7 Hz, COC**H**$_2$), 2.34 (dd, 1H, *J* = 16.6, 7.8 Hz, COC**H**$_2$), 1.98 (s, 3H, NHCOC**H**$_3$), 1.94 (m, 1H, C**H**), 1.75-1.64 (m, 4H, CHClC**H**$_2$) 1.56-1.15 (m, 10H, C**H**$_2$), 0.91 (m, 6H, C**H**$_3$). **Enol form:** δ 12.60 (s, 1H, CHCO**H**), 5.92 (s, 1H, N**H**), 5.45 (s, 1H, C**H**COH), 3.89 (m, 1H, C**H**Cl), 3.50-3.43 (m, 2H, NHC**H**$_2$), 3.12-3.07 (m, 2H, SC**H**$_2$), 2.18 (dd, 1H, *J* = 7.8, 2.3 Hz, COC**H**$_2$), 2.02 (m, 1H, COC**H**$_2$), 1.98 (s, 3H, NHCOC**H**$_3$), 1.94 (m, 1H, C**H**), 1.75-1.64 (m, 4H, CHClC**H**$_2$) 1.56-1.15 (m, 10H, C**H**$_2$), 0.91 (m, 6H,

C**H**$_3$). ^{13}C-NMR (125 MHz; CDCl3): δ 202.2, 194.4, 192.5, 177.0, 170.6, 170.5, 100.55, 64.5, 57.9, 51.0 (2), 42.8, 42.7, 40.1, 39.5, 38.7, 38.5, 36.8, 31.3, 29.9, 29.5, 29.1, 28.9, 28.1, 26.8, 26.7, 23.5, 23.4, 22.5, 19.9, 19.8, 14.2.

4.5. References

(1) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P. Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J. Am. Chem. Soc.* **2012**, *134*, 18518.

(2) Sydor, P. K.; Barry, S. M.; Odulate, O. M.; Barona-Gomez, F.; Haynes, S. W.; Corre, C.; Song, L. J.; Challis, G. L. Regio- and stereodivergent antibiotic oxidative carbocyclizations catalysed by Rieske oxygenase-like enzymes. *Nat. Chem.* **2011**, *3*, 388.

(3) Cooley, R. B.; Rhoads, T. W.; Arp, D. J.; Karplus, P. A. A diiron protein autogenerates a valine-phenylalanine cross-link. *Science* **2011**, *332*, 929.

(4) Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat. Chem.* **2015**, *7*, 431.

(5) Barry, S. M.; Challis, G. L. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal.* **2013**, *3*, 2362.

(6) Withall, D. M.; Haynes, S. W.; Challis, G. L. Stereochemistry and mechanism of undecylprodigiosin oxidative carbocyclization to streptorubin B by the Rieske oxygenase RedG. *J. Am. Chem. Soc.* **2015**, *137*, 7889.

(7) Salem, S. M.; Kancharla, P.; Florova, G.; Gupta, S.; Lu, W. L.; Reynolds, K. A. Elucidation of final steps of the marineosins biosynthetic pathway through identification and characterization of the corresponding gene cluster. *J. Am. Chem. Soc.* **2014**, *136*, 4565.

(8) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical *S*-adenosylmethionine enzymes. *Chem. Rev.* **2014**, *114*, 4229.

(9) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **2005**, *436*, 1191.

(10) Kelly, W. L.; Boyne, M. T.; Yeh, E.; Vosburg, D. A.; Galonic, D. P.; Kelleher, N. L.; Walsh, C. T. Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. *Biochemistry* **2007**, *46*, 359.

(11) Gu, L. C.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Hakansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459*, 731.

(12) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E. Structures of cylindrocyclophanes A-F. *Tetrahedron* **1992**, *48*, 3001.

(13) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. Cylindrocyclophanes with proteasome inhibitory activity from the cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2010**, *73*, 1529.

(14) Kang, H. S.; Santarsiero, B. D.; Kim, H.; Krunic, A.; Shen, Q.; Swanson, S. M.; Chai, H.; Kinghorn, A. D.; Orjala, J. Merocyclophanes A and B, antiproliferative cyclophanes from the cultured terrestrial cyanobacterium *Nostoc* sp. *Phytochemistry* **2012**, *79*, 109.

(15) Luo, S. W.; Kang, H. S.; Krunic, A.; Chlipala, G. E.; Cai, G. P.; Chen, W. L.; Franzblau, S. G.; Swanson, S. M.; Orjala, J. Carbamidocyclophanes F and G with anti-*Mycobacterium tuberculosis* activity from the cultured freshwater cyanobacterium *Nostoc* sp. *Tetrahedron Lett.* **2014**, *55*, 686.

(16) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. Carbamidocyclophanes A-E, chlorinated paracyclophanes with cytotoxic and antibiotic activity from the Vietnamese cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2007**, *70*, 499.

(17) Preisitsch, M.; Harmrolfs, K.; Pham, H. T. L.; Heiden, S. E.; Fuessel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H. J.; Muller, R.; Mundt, S. Anti-MRSA-acting carbamidocyclophanes H-L from the Vietnamese cyanobacterium *Nostoc* sp. CAVN2. *J. Antibiot.* **2015**, *68*, 600.

(18) Cserzo, M.; Wallin, E.; Simon, I.; vonHeijne, G.; Elofsson, A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: The dense alignment surface method. *Protein Eng.* **1997**, *10*, 673.

(19) Pasquier, C.; Promponas, V. J.; Palaios, G. A.; Hamodrakas, J. S.; Hamodrakas, S. J. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the Swiss-Prot database: The PRED-TMR algorithm. *Protein Eng.* **1999**, *12*, 381.

(20) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567.

(21) Käll, L.; Krogh, A.; Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **2007**, *35*, W429.

(22) Soding, J.; Biegert, A.; Lupas, A. N. The HHPred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244.

(23) Choi, Y. S.; Zhang, H. J.; Brunzelle, J. S.; Nair, S. K.; Zhao, H. M. In vitro reconstitution and crystal structure of *p*-aminobenzoate *N*-oxygenase (AurF) involved in aureothin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6858.

(24) Nordlund, P.; Eklund, H. Di-iron-carboxylate proteins. *Curr. Opin. Struct. Biol.* **1995**, *5*, 758.

(25) Krebs, C.; Bollinger, J. M.; Booker, S. J. Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like 'di-iron-carboxylate' proteins. *Curr. Opin. Chem. Biol.* **2011**, *15*, 291.

(26) Fox, B. G.; Lyle, K. S.; Rogge, C. E. Reactions of the diiron enzyme stearoyl-acyl carrier protein desaturase. *Acc. Chem. Res.* **2004**, *37*, 421.

(27) Shanklin, J.; Guy, J. E.; Mishra, G.; Lindqvist, Y. Desaturases: Emerging models for understanding functional diversification of diiron-containing enzymes. *J. Biol. Chem.* **2009**, *284*, 18559.

(28) Tinberg, C. E.; Lippard, S. J. Dioxygen activation in soluble methane monooxygenase. *Acc. Chem. Res.* **2011**, *44*, 280.

(29) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **2006**, *Chapter 5*, Unit 5.6.

(30) Welker, M.; von Dohren, H. Cyanobacterial peptides—nature's own combinatorial biosynthesis. *FEMS Microbiol. Rev.* **2006**, *30*, 530.

(31) Van Wagoner, R. M.; Drummond, A. K.; Wright, J. L. Biogenetic diversity of cyanobacterial metabolites. *Adv. Appl. Microbiol.* **2007**, *61*, 89.

(32) Okino, T.; Matsuda, H.; Murakami, M.; Yamaguchi, K. Microginin, an angiotensin-converting enzyme-inhibitor from the blue-green-alga *Microcystis aeruginosa*. *Tetrahedron Lett.* **1993**, *34*, 501.

(33) Kodani, S.; Suzuki, S.; Ishida, K.; Murakami, M. Five new cyanobacterial peptides from water bloom materials of lake Teganuma (Japan). *FEMS Microbiol. Lett.* **1999**, *178*, 343.

(34) Ishida, K.; Matsuda, H.; Murakami, M. Four new microginins, linear peptides from the cyanobacterium *Microcystis aeruginosa*. *Tetrahedron* **1998**, *54*, 13475.

(35) Kramer, D. (Cyano Biotech GmbH, Germany; Univ. Berlin Humboldt, Germany). Microginin producing proteins and nucleic acids encoding a microginin gene cluster as well as methods for creating novel microginins. *US Patent* 7,846,686B2, Dec. 7, **2010.**

(36) Leão, P. N.; Nakamura, H.; Costa, M.; Pereira, A. R.; Martins, R.; Vasconcelos, V.; Gerwick, W. H.; Balskus, E. P. Biosynthesis-assisted structural elucidation of the bartolosides, chlorinated aromatic glycolipids from cyanobacteria. *Angew. Chem. Int. Ed.* **2015**, *54*, 11063.

(37) Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H. Combining mass spectrometric metabolic profiling with genomic analysis: A powerful approach for discovering natural products from cyanobacteria. *J. Nat. Prod.* **2015**, *78*, 1671.

(38) Preisitsch, M.; Niedermeyer, T. H. J.; Heiden, S. E.; Neidhardt, I.; Kumpfmuller, J.; Wurster, M.; Harmrolfs, K.; Wiesner, C.; Enke, H.; Muller, R.; Mundt, S. Cylindrofridins A-C, linear cylindrocyclophane-related alkylresorcinols from the cyanobacterium *Cylindrospermum stagnale*. *J. Nat. Prod.* **2016**, *79*, 106.

(39) Schwartz, R.; Ting, C. S.; King, J. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res.* **2001**, *11*, 703.

(40) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of e-coli. *Science* **2001**, *291*, 1790.

(41) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* **1996**, *3*, 923.

(42) Kennedy, M. C.; Kent, T. A.; Emptage, M.; Merkle, H.; Beinert, H.; Münck, E. Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase. *J. Biol. Chem.* **1984**, *259*, 4463.

(43) Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T. Nature's inventory of halogenation catalysts: Oxidative strategies predominate. *Chem. Rev.* **2006**, *106*, 3364.

(44) Bollinger, J. M.; Price, J. C.; Hoffart, L. M.; Barr, E. W.; Krebs, C. Mechanism of taurine: Alpha-ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Eur. J. Inorg. Chem.* **2005**, 4245.

(45) Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L. Dioxygen activation at mononuclear non-heme iron active sites: Enzymes, models, and intermediates. *Chem. Rev.* **2004**, *104*, 939.

(46) Neumann, C. S.; Fujimori, D. G.; Walsh, C. T. Halogenation strategies in natural product biosynthesis. *Chem. Biol.* **2008**, *15*, 99.

(47) Banerjee, R.; Proshlyakov, Y.; Lipscomb, J. D.; Proshlyakov, D. A. Structure of the key species in the enzymatic oxidation of methane to methanol. *Nature* **2015**, *518*, 431.

(48) Yeh, E.; Garneau, S.; Walsh, C. T. Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3960.

(49) Vaillancourt, F. H.; Yin, J.; Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a non-heme Fe-II alpha-ketoglutarate and O_2 dependent halogenase. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10111.

(50) Hillwig, M. L.; Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **2014**, *10*, 921.

(51) Nakamura, H.; Wang, J. X.; Balskus, E. P. Assembly line termination in cylindrocyclophane biosynthesis: Discovery of an editing type II thioesterase domain in a type I polyketide synthase. *Chem. Sci.* **2015**, *6*, 3816.

(52) Eustaquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **2008**, *4*, 69.

(53) Wagner, C.; El Omari, M.; Konig, G. M. Biohalogenation: Nature's way to synthesize halogenated metabolites. *J. Nat. Prod.* **2009**, *72*, 540.

(54) Preisitsch, M.; Heiden, S. E.; Beerbaum, M.; Niedermeyer, T. H.; Schneefeld, M.; Herrmann, J.; Kumpfmuller, J.; Thurmer, A.; Neidhardt, I.; Wiesner, C.; Daniel, R.; Müller, R.; Bange, F. C.; Schmieder, P.; Schweder, T.; Mundt, S. Effects of halide ions on the carbamidocyclophane biosynthesis in *Nostoc* sp. CAVN2. *Mar. Drugs* **2016**, *14*.

(55) Gribble, G. W. Natural organohalogens: A new frontier for medicinal agents? *J. Chem. Educ.* **2004**, *81*, 1441.

(56) Harris, C. M.; Kannan, R.; Kopecka, H.; Harris, T. M. The role of the chlorine substituents in the antibiotic vancomycin—preparation and characterization of monodechlorovancomycin and didechlorovancomycin. *J. Am. Chem. Soc.* **1985**, *107*, 6652.

(57) Groll, M.; Huber, R.; Potts, B. C. M. Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of beta-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **2006**, *128*, 5136.

(58) Pereira, E. R.; Belin, L.; Sancelme, M.; Prudhomme, M.; Ollier, M.; Rapp, M.; Severe, D.; Riou, J. F.; Fabbro, D.; Meyer, T. Structure-activity relationships in a series of substituted indolocarbazoles: Topoisomerase I and protein kinase C inhibition and antitumoral and antimicrobial properties. *J. Med. Chem.* **1996**, *39*, 4471.

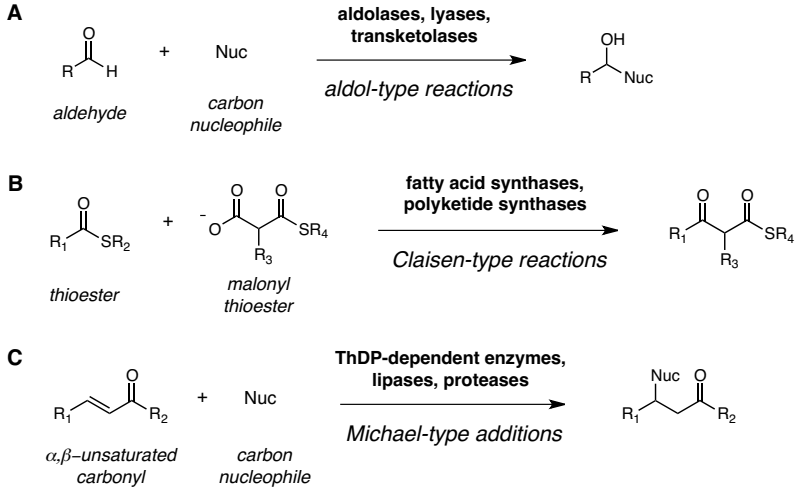
(59) Holmquist, H. E.; Rothrock, H. S.; Theobald, C. W.; Englund, B. E. Some decomposition and rearrangement products of decahydronaphthalene hydroperoxide. *J. Am. Chem. Soc.* **1956**, *78*, 5339.

(60) Mino, T.; Masuda, S.; Nishio, M.; Yamashita, M. Synthesis of lactones by Baeyer–Villiger oxidation with magnesium monoperphthalate hexahydrate. *J. Org. Chem.* **1997**, *62*, 2633.

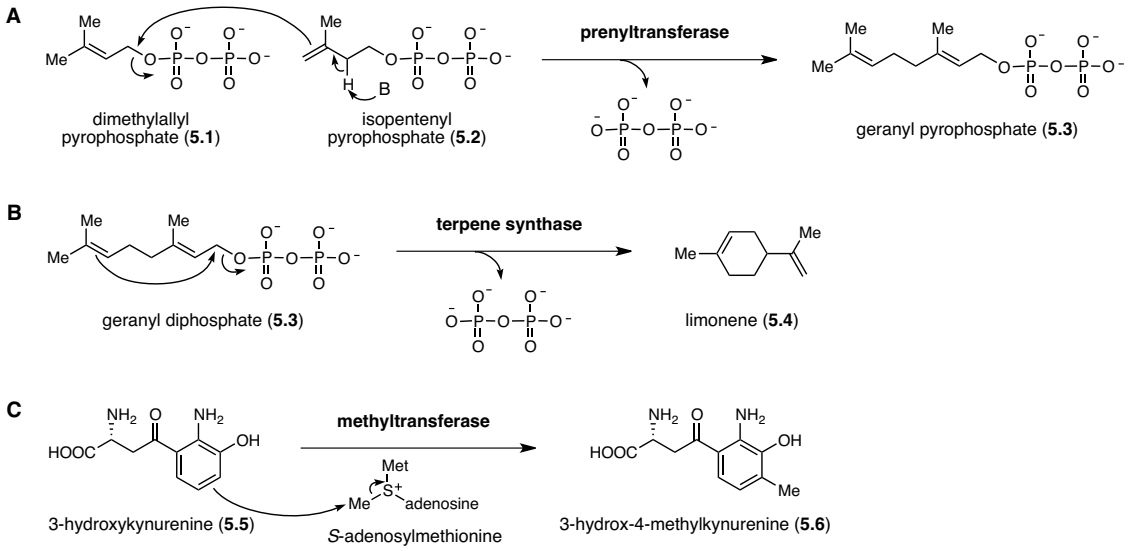
Chapter 5. Complete elucidation of the cylindrocyclophane biosynthesis: Discovery of the novel Carbon–Carbon bond-forming enzyme

5.1. Introduction

Carbon–carbon bond formation is essential for building organic molecules in both biology and synthetic chemistry. While many enzymes form C–C bonds, most utilize carbonyl electrophiles to catalyze aldol-type (**Scheme 5.1A**) or Claisen-type (**Scheme 5.1B**) 1,2-additions.[1,2] Some enzymes are also capable of catalyzing Michael-type 1,4-additions (**Scheme 5.1C**).[3-5] Other common enzymatic C–C bond formations include addition of isopentenyl pyrophosphates (**5.2**) or other larger pyrophosphates by prenyltransferases (**Scheme 5.2A**),[6] cyclizations catalyzed by terpene synthases (**Scheme 5.2B**),[7,8] and methylation reactions catalyzed by *S*-adenosylmethionine dependent methyltransferases (**Scheme 5.2C**).[9] With growing interest in using enzymes in industrial processes, there is a need for the discovery of new C–C bond-forming enzymes that may be developed into useful biocatalysts.[10-12]

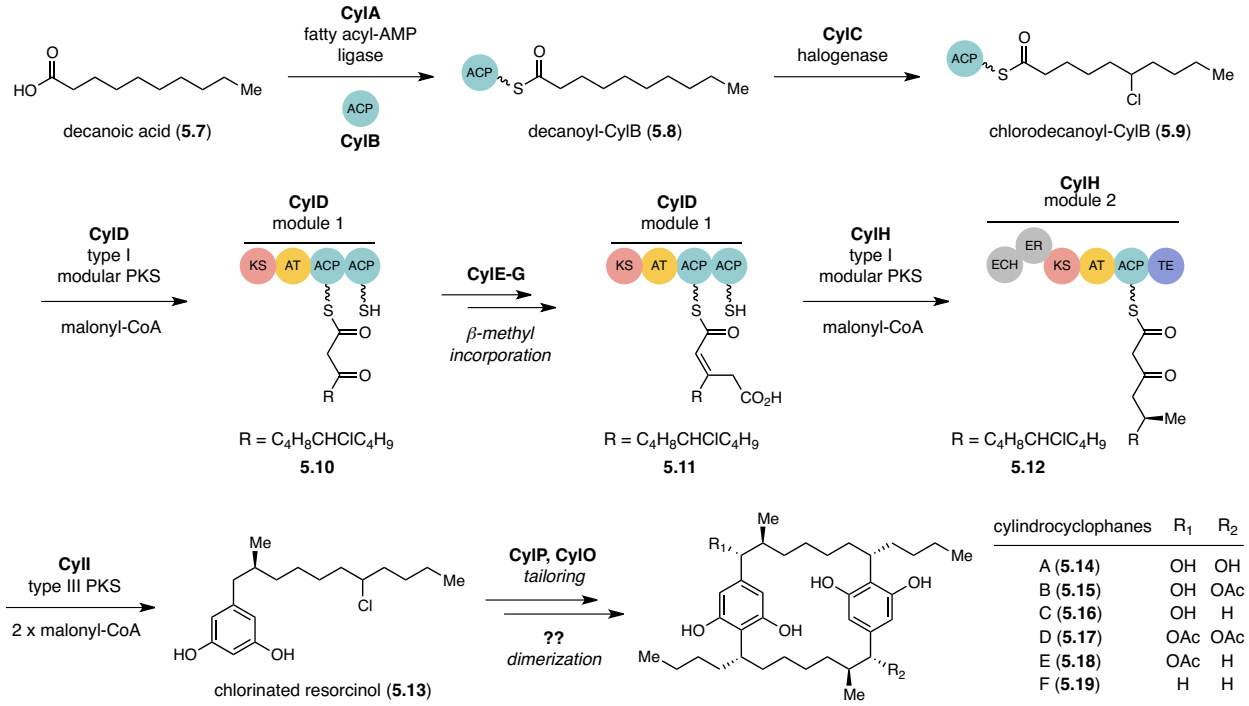


Scheme 5.1. Major types of enzyme catalyzed C–C bond formations involving carbonyl electrophiles. **A**) Aldol-type 1,2-additions to aldehydes. **B**) Claisen-type 1,2-additions to thioesters. **C**) Michael-type 1,4-addition to □,□-unsaturated carbonyls.



Scheme 5.2: Examples of enzymatic C–C bond formations involving non-carbonyl electrophiles. **A**) Condensation of isoprene units by prenyltransferases. **B**) Cyclization catalyzed by terpene synthase. **C**) Methylation catalyzed by SAM-dependent methyltransferase.

One of our major goals for studying the biosynthesis of the cylindrocyclophanes was to identify the enzyme that catalyzes the C–C bond formation in the dimerization step (**5.13** to **5.19**, **Scheme 5.3**). Based on our discovery of the novel halogenase CylC and the *in vivo* characterization of its function to generate chlorinated decanoyl-CylB (**5.9**), we hypothesized that the cylindrocyclophane biosynthesis proceeds through cryptic chlorination. This revised biosynthetic hypothesis was supported by the results of the substrate scope assays for the fatty acid activating enzyme CylA, the type I PKS CylH, and the type III PKS CylI. CylA selectively activates unfunctionalized decanoic acid (**5.7**) and does not activate the pre-chlorinated decanoic acid derivative. On the other hand, both CylH and CylI are capable of accepting chlorinated substrates. We therefore hypothesize that the final paracyclophane-forming, dimerization step requires intermolecular C–C bond formations between the nucleophilic resorcinol and the electrophilic alkyl chloride of the chlorinated resorcinol derivative (**5.13**). To the best of our knowledge, an enzymatic C–C bond formation that involves a secondary alkyl chloride has never been reported.



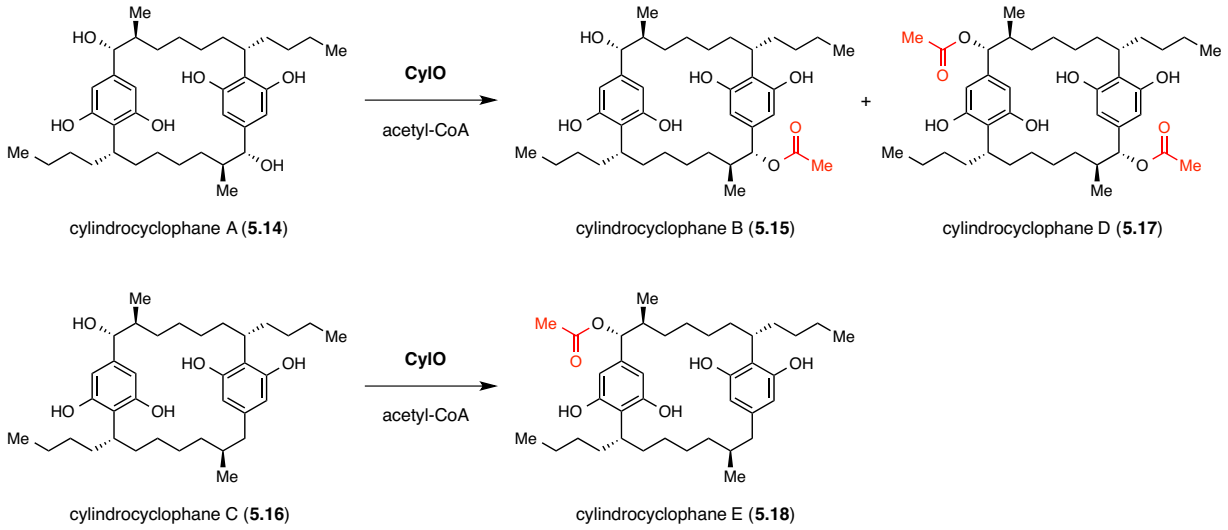
Scheme 5.3: Revised biosynthetic hypothesis for cylindrocyclophane biosynthesis involving cryptic chlorination.

Our previous work shows that the *in situ* generation of the chlorinated resorcinol (**5.13**) does not lead to spontaneous dimerization, which indicates that another enzyme is required for the dimerization step (see **Chapter 4.2.5**). The *cyl* gene cluster in *C. licheniforme* ATCC 29412 contained six possible tailoring genes (*cylJ-O*) that we had not assigned functions to (**Table 5.1**). Based on BLAST and conserved domain search results, the six genes were annotated as following: *cylJ* encodes a methyltransferase, *cylK* encodes a hemolysin-type calcium binding protein, *cylL* encodes a flavin-dependent oxidoreductase, *cylM* encodes an oxidoreductase, *cylN* encodes a glycosyltransferase and *cylO* encodes an acetyltransferase.

Table 5.1: Annotated functions of the tailoring proteins in the *cyl* gene cluster

Protein	Size, aa	Proposed function	Protein homolog	Accession number	Identity/ similarity %
CylJ	181	Phospholipid methyltransferase	PCC7424_2732, *Cyanothece sp.* PCC 7424	ACK71144	61/80
CylK	651	Hemolysin-type calcium-binding region	FJSC11DRAFT_3960, *Fischerella sp.* JSC-11	EHC09323	45/63
CylL	535	Flavin-dependent oxidoreductase	*Thiorhodovibrio* sp. 970	ZP_09809663	36/57
CylM	232	Aldo/keto reductase	*Mycobacterium xenopi* RVm700367	ZP_00979622	35/79
CylN	435	Glycosyltransferase_MGT family	*Calothrix sp.* PCC 7507	YP_007065913	59/97
CylO	448	Transferase family	*Microcystis aeruginosa* PCC 9443	WP_002766670	58/77

The function of one of the tailoring enzymes, CylO, is relatively easy to predict. CylO is an acetyltransferase that is expected to catalyze acetylation of benzylic hydroxyl groups to produce cylindrocyclophanes B (**5.15**), D (**5.17**) and E (**5.18**, see **Scheme 5.4**). There is also a possibility that CylO catalyzes acetylation of monomeric resorcinol units prior to the dimerization step. Another tailoring enzyme, CylN, is annotated as a glycosyltransferase. While glycosylated cylindrocyclophanes have not been isolated, glycosylation has been observed for the related natural products, nostocyclophanes A and B (**Figure 5.1**).[13] We envisage that the cylindrocyclophane biosynthetic pathway in *C. licheniforme* ATCC 29412 may also produce analogous glycosylated cylindrocyclophanes under certain conditions that have not yet been identified.



Scheme 5.4: Predicted acetyltransferase activity of CylO to generate cylindrocyclophanes B, D and E.

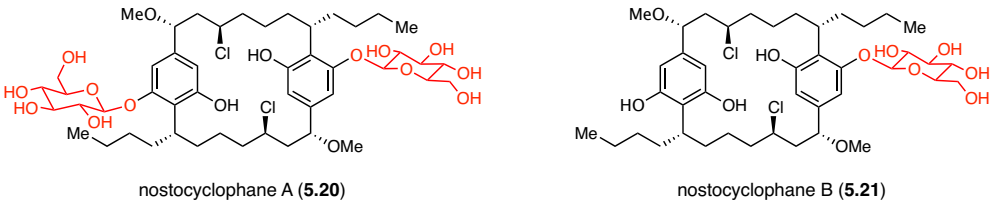
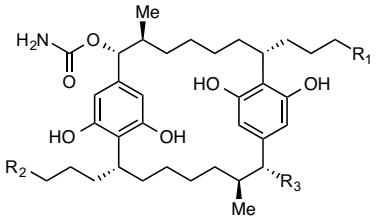


Figure 5.1: Structures of glycosylated nostocyclophanes A and B.

The enzyme required for the construction of the paracyclophane scaffold is likely conserved in all the biosynthetic gene clusters of the cylindrocyclophanes (**5.14**-**5.19**) and their carbamoylated analogs, the carbamidocyclophanes (**5.22**-**5.32**, see **Figure 5.2**). To identify the enzyme catalyzing the final dimerization from the other possible tailoring enzymes CylJ-M, we compared the gene organizations of the cylindrocyclophane (*cyl*) biosynthetic gene clusters in *C. licheniforme* ATCC 29412,[14] *C. stagnale* PCC 7417,[15] and *Nostoc* sp. UIC 10022A,[16] as well as the carbamidocyclophane (*cab*)[17] biosynthetic gene cluster in *Nostoc* sp. CAVN2 (**Figure 5.3**).[17]



carbamidocyclophanes	R_1	R_2	R_3
A (**5.22**)	$CHCl_2$	$CHCl_2$	$OCONH_2$
B (**5.23**)	$CHCl_2$	CH_2Cl	$OCONH_2$
C (**5.24**)	$CHCl_2$	CH_3	$OCONH_2$
D (**5.25**)	CH_2Cl	CH_3	$OCONH_2$
E (**5.26**)	CH_3	CH_3	$OCONH_2$
F (**5.27**)	$CHCl_2$	$CHCl_2$	OH
H (**5.28**)	CH_3	CH_3	OH
I (**5.29**)	CH_2Cl	CH_3	OH
J (**5.30**)	CH_2Cl	CH_2Cl	$OCONH_2$
K (**5.31**)	$CHCl_2$	CH_3	OH
L (**5.32**)	$CHCl_2$	CH_2Cl	OH

Figure 5.2: Structures of the carbamidocyclophanes A-F and H-L isolated from *Nostoc* sp. CAVN2.

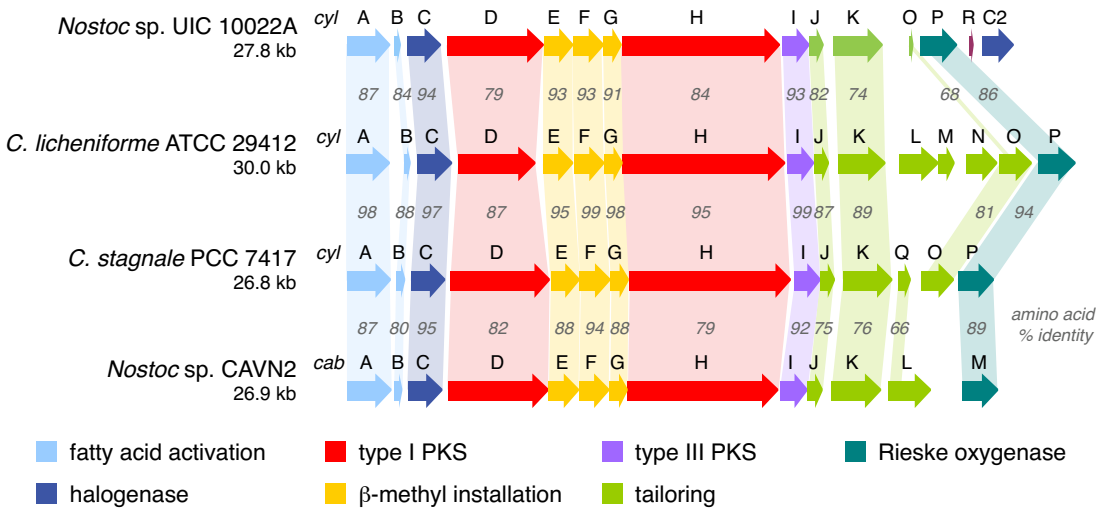


Figure 5.3: Comparison of the open reading frame (ORF) diagrams of the four sequenced *cyl* and *cab* biosynthetic gene clusters. The conserved genes are highlighted and are labeled with % amino acid identity.

Comparison of the four gene clusters revealed that three of the possible tailoring genes (*cylL-N*) in the *C. licheniforme* gene cluster are entirely absent in the other three gene clusters. The gene encoding the acetyltransferase *cylO* appears to be only present in the *C. licheniforme* and the *C. stagnale* gene clusters. This observation is consistent with the isolation of acetylated cylindrocyclophanes B (**5.15**) and D (**5.17**) from *C. stagnale* PCC 7417 (also cross-listed as *Cylindrospermum* sp. ATCC 29204).[18] *Nostoc* sp. UIC 10022A only has a partial *cylO* and *Nostoc* sp. CAVN2 lacks a *cylO* homolog entirely, and acetylated cylindrocyclophanes and carbamidocyclophanes have not been isolated from these strains.[16,17,19] *Nostoc* sp. CAVN2 strain instead has *cabL*, which encodes a carbamoyltransferase that is predicted to install the benzylic carbamoyl groups.

Out of the possible tailoring genes in the three *cyl* and the *cab* gene clusters, only *cylJ* and *cylK* are conserved. Based on this comparative analysis of the four gene clusters, we hypothesized that either CylJ or CylK is responsible for catalyzing the C–C bond formation in the final dimerization step of the cylindrocyclophane biosynthesis. This work describes our efforts to characterize the functions of CylJ and CylK, which led to the discovery of a novel C–C bond-forming enzyme that catalyzes a Friedel–Crafts alkylation.

5.2. Results and Discussions

5.2.1. Investigation of the function of CylJ

CylJ is annotated as a SAM-dependent methyltransferase, and prediction of membrane protein topology using online tools, such as TMHMM[20] and PRED-TMR,[21] indicate that this enzyme is a transmembrane protein with five predicted transmembrane helices. Using the HHPred search,[22] CylJ was identified to have a secondary structural homology to an integral membrane isoprenylcysteine carboxyl methyltransferase (ICMT).[23] ICMT is known to catalyze posttranslational methylation of proteins that have a C-terminal CAAX motif with prenylation on the cysteine residue.[23] A homology model of CylJ was generated using the crystal structure of ICMT from *Methanosarcina acetivorans* as a template (**Figure 5.4**).

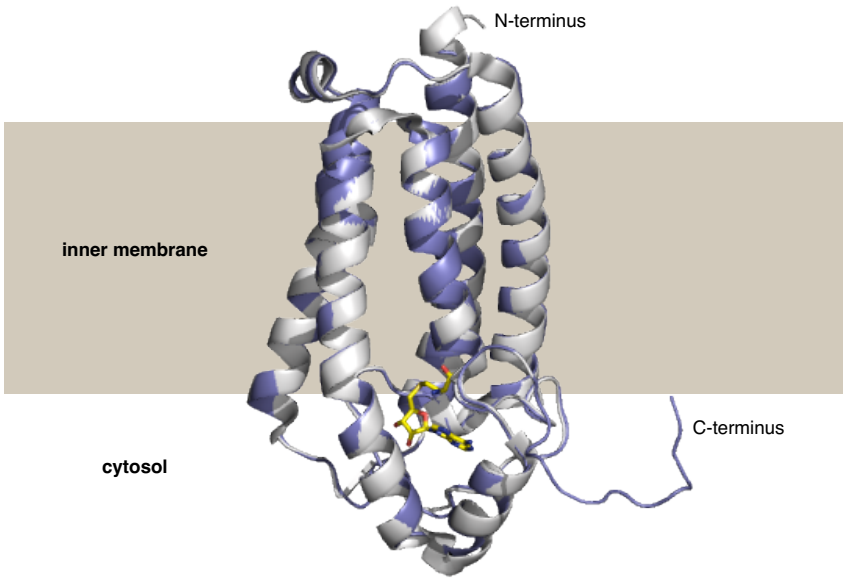


Figure 5.4: A homology model of CylJ (blue) overlaid with the structure of ICMT (white). *S*-adenosylhomocysteine (yellow) co-crystallized with ICMT binds to the cytosolic region of the protein.

While we predict that CylJ is an integral membrane methyltransferase with homology to ICMT, its function in cylindrocyclophane biosynthesis is not apparent. To obtain more information on CylJ, we examined the genomic context of CylJ homologs in other biosynthetic gene clusters. The analysis revealed that CylJ homologs are often found adjacent to type III PKSs in multiple biosynthetic gene clusters. The clustering of CylJ and CylI homologs in multiple organisms indicate that these enzymes might function together or consecutively.

The functions of several CylJ homologs that are found adjacent to type III PKSs have previously been characterized.[24-26] For example, the *srs* gene cluster in *Streptomyces griseus* encodes a type III PKS (*srsA*), a methyltransferase (*srsB*) and a flavoprotein hydroxylase (*srsC*), and this gene cluster has been shown to produce methoxylated resorcinols (**Figure 5.5**).[25,26] These resorcinols are hypothesized to integrate into the bacterial cell membrane, which increase the rigidity of the membrane to allow the growth of *S. griseus* in the presence of β-lactam antibiotics.[26] The CylJ homolog, SrsB, has been shown to catalyze *O*-methylation of resorcylic acid (**5.35**) using *S*-adenosylmethionine (SAM) as the methyl donor (**Reaction 2**, **Figure 5.5B**).[25] Thus, we hypothesize that CylJ might catalyze similar *O*-methylation of the

monomeric resorcinol. However, methoxylated cylindrocyclophanes or resorcinols have not been isolated from the cylindrocyclophane-producing strains.

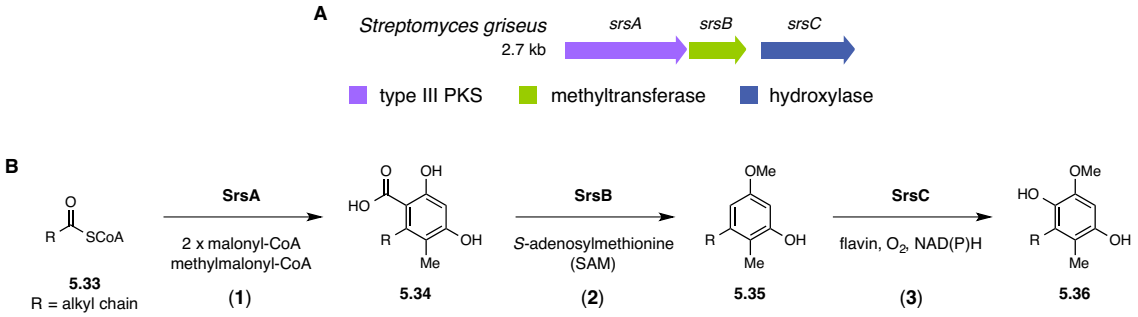


Figure 5.5: The production of alkylresorcinol methyl ethers by the *srs* gene cluster.

One potential role of CylJ in the cylindrocyclophane biosynthesis is detoxification. Previous feeding experiments in the native producer, *C. licheniforme* ATCC 29412 (discussed in **Chapter 2**), revealed that the addition of the synthetic resorcinol (**5.37**) to the culture results in cytotoxicity (**Figure 5.6**).[14] The LC-HRMS analysis of the culture extracts also showed that the non-chlorinated monomeric resorcinol (**5.38**, see **Scheme 5.5**) does not accumulate under normal culturing conditions. However, over-activation of the cylindrocyclophane biosynthetic pathway might lead to accumulation of the cytotoxic non-chlorinated resorcinol (**5.38**). In cases the non-chlorinated resorcinol (**5.38**) accumulates, CylJ may catalyze *O*-methylation to form methoxylated resorcinol (**5.39**), which may have reduced cytotoxicity (**Scheme 5.5**). Results of the deuterium-labeled decanoic acid feeding experiments (see **Chapter 2**) support this hypothesis: We observed masses corresponding to d_4- (**5.42**) or d_{19}-methoxylated resorcinol (**5.44**) in the organic extracts of *C. licheniforme* ATCC 29412 fed with d_4-(*R*)-decanoic acid (**5.40**), d_4-(*S*)-decanoic acid (**5.41**) or d_{19}- decanoic acid (**5.43**, see **Figure 5.7**). The addition of deuterated decanoic acids to the cultures of *C. licheniforme* may have resulted in inefficient chlorination of the pathway intermediates, possibly due to kinetic isotope effects, leading to the overproduction of the non-chlorinated resorcinol (**5.38**). In situations like this, CylJ may catalyze *O*-methylation to form the methoxylated resorcinol (**5.44**), preventing the accumulation of the cytotoxic non-chlorinated resorcinol (**5.38**) as a detoxification strategy in the native producer**.**

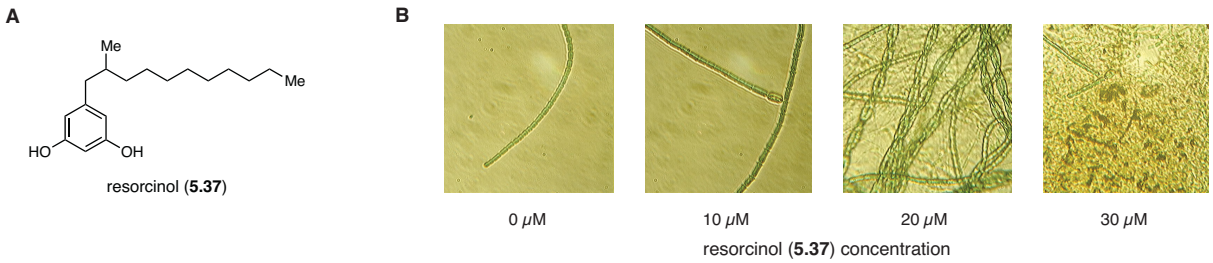
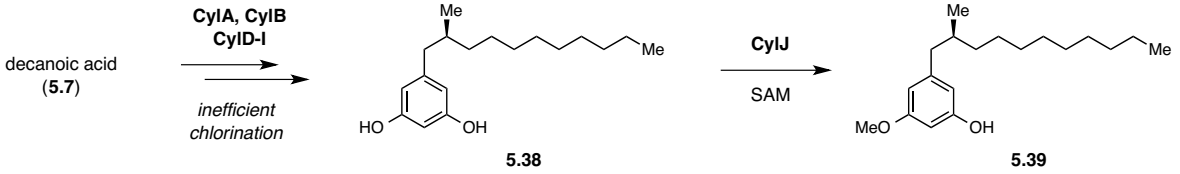


Figure 5.6: Addition of synthetic resorcinol (**5.37**) to the cultures of *C. licheniforme* ATCC 29412 leads to cell lysis.



Scheme 5.5: Possible role of CylJ in detoxification through *O*-methylation of the unfunctionalized resorcinol (**5.38**).

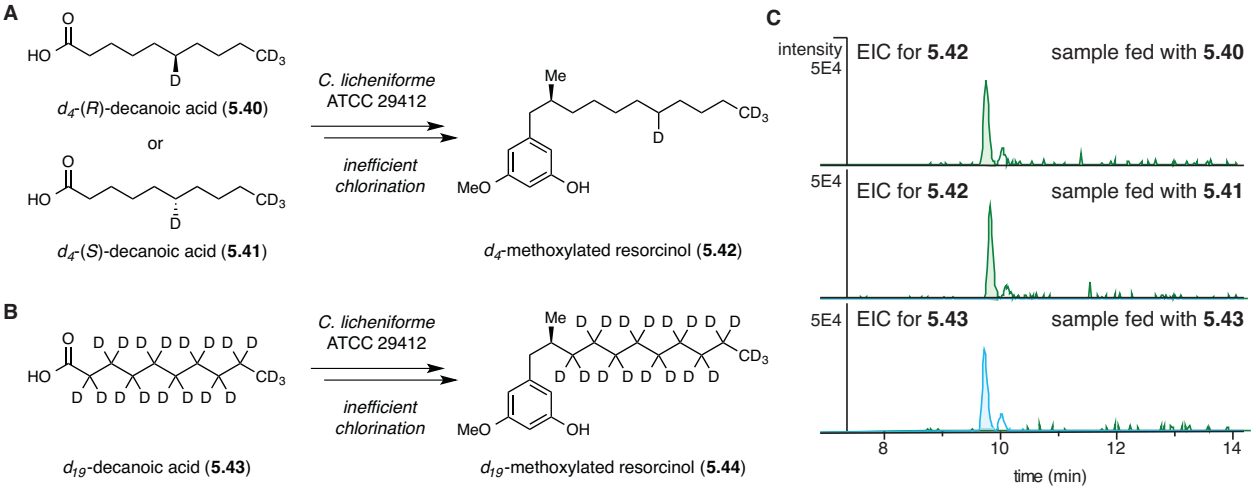


Figure 5.7: Feeding of deuterated decanoic acid to *C. licheniforme* ATCC 29412 results in the production of deuterium-labeled methoxylated resorcinol. **A**) Formation of d_4-methoxylated resorcinol. **B**) Formation of d_{19}-methoxylated resorcinol. **C**) Extracted ion chromatograms of d_4-and d_{19}-methoxylated resorcinol.

While we have cloned *cylJ* into various *E. coli* expression hosts for heterologous expression, we have not been able to successfully express CylJ. This is partly because CylJ is a transmembrane protein and the level of expression is expected to be relatively low. Based on the functions of the CylJ homologs and the observations made in the feeding studies, CylJ is predicted to catalyze the detoxification of resorcinol and does not catalyze the dimerization step.

5.2.2. Bioinformatic analyses of CylK

The only other conserved gene in all the *cyl* and the *cab* gene clusters is *cylK*. CylK is annotated as a hemolysin-type calcium-binding protein with β-propeller repeats, but has low primary sequence homology to any characterized proteins. A NCBI conserved domain search[27] and a HHPred[22] search for secondary structural homology revealed that CylK is likely a di-domain protein with an N-terminal calcium-binding domain and a C-terminal domain containing NHL repeats (**Figure 5.8**). The HHPred search showed that the N-terminal domain of CylK has homology to proteins containing a repeat-in-toxin (RTX) motif. The RTX motif, which consists of a glycine- and aspartate-rich sequence that binds to Ca^{2+} ions,[28] was first described in a group of extracellular proteins that exhibit a cytotoxic pore-forming activity.[29-31] This motif is commonly found at the C-terminus of proteins and serves as a non-cleavable secretion signal for extracellular transport by the type I secretion system in Gram-negative bacteria.[28] Unlike other proteins with the RTX motif, CylK has the RTX calcium-binding motif at the N-terminus instead of C-terminus, which might not function as a secretion signal.

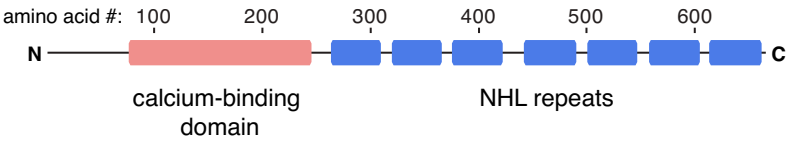


Figure 5.8: Domains and repeats found in CylK.

The NHL repeat was first identified in three proteins, NCL-1, HT2A and LIN-41 (abbreviated as NHL), and typically consists of a 40-50 residue sequence that is rich in glycine and hydrophobic residues as well as a cluster of charged residues (such as Asp, Asn, His, and Arg) near its C-terminal end.[32] The NHL repeats in the C-terminal domain of CylK are predicted to form a β-propeller fold with 7-blades, each of which consists of a four-stranded antiparallel β-sheet motif (**Figure 5.9A**).[33] Domains with the β-propeller fold typically take a disc shape with a central tunnel lined with hydrogen donors and acceptors (**Figure 5.9B**).[34] Many of the known β-propeller proteins use their central tunnel, specifically the narrower entry site, to coordinate a ligand or to catalyze reactions.[33]

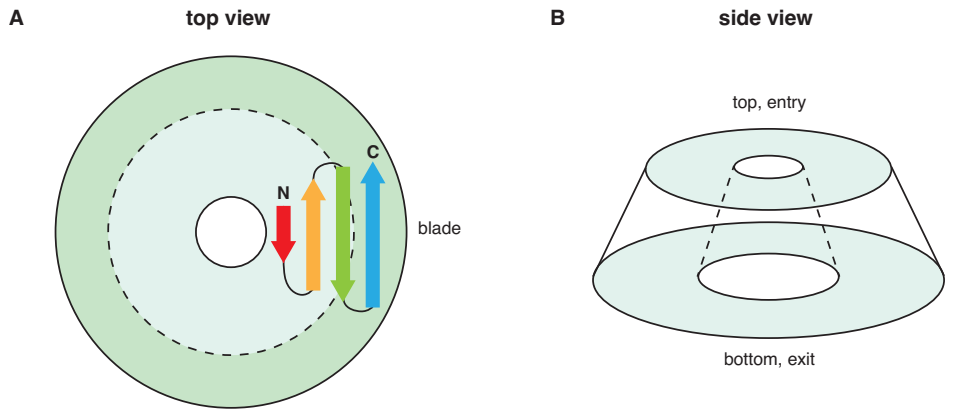


Figure 5.9: Diagram of a typical domain with β-propeller fold. **A**) Domains with a β-propeller fold has multiple blades that consist of a four-stranded antiparallel β-sheets. **B**) β-propeller protein forms a disc shape with a central tunnel.

Examples of structurally characterized enzymes with homology to CylK include the R-modules of the C-5 alginate epimerases AlgE4/6[35] for the N-terminal calcium-binding domain and virginiamycin B lyase[36] for the C-terminal NHL repeat domain. A homology model of CylK was generated using the structures of AlgE6 R-module and Vgb lyase as templates (**Figure 5.10**). C-5 alginate epimerases are periplasmic or extracellular enzymes that catalyze epimerization of β-D-mannuronic acid in polymannuronate (**5.45**, see **Scheme 5.6**) to α-L-guluronic acid in the biosynthesis of alginates (**5.46**), which are unbranched biopolymers produced by brown algae and bacteria.[37-39] The regulatory R-modules of alginate epimerases require calcium for activity as well as for structural integrity and these modules are thought to be important for controlling the catalytic activity of the A-modules.[35] In addition, the R-modules are also important for tight substrate binding for the alginate epimerases to processively catalyze multiple rounds of epimerization on each polymeric substrate.[40,41] The Vgb lyase is responsible for catalyzing the C–O bond cleavage of streptogramin antibiotics, such as quinupristin (**5.47**, see **Scheme 5.7**), as one of the strategies bacteria employs to confer antibiotic resistance.[42] The Vgb lyase has a seven-bladed β-propeller fold and requires a divalent metal ion such as Mg^{2+} for catalysis.[36]

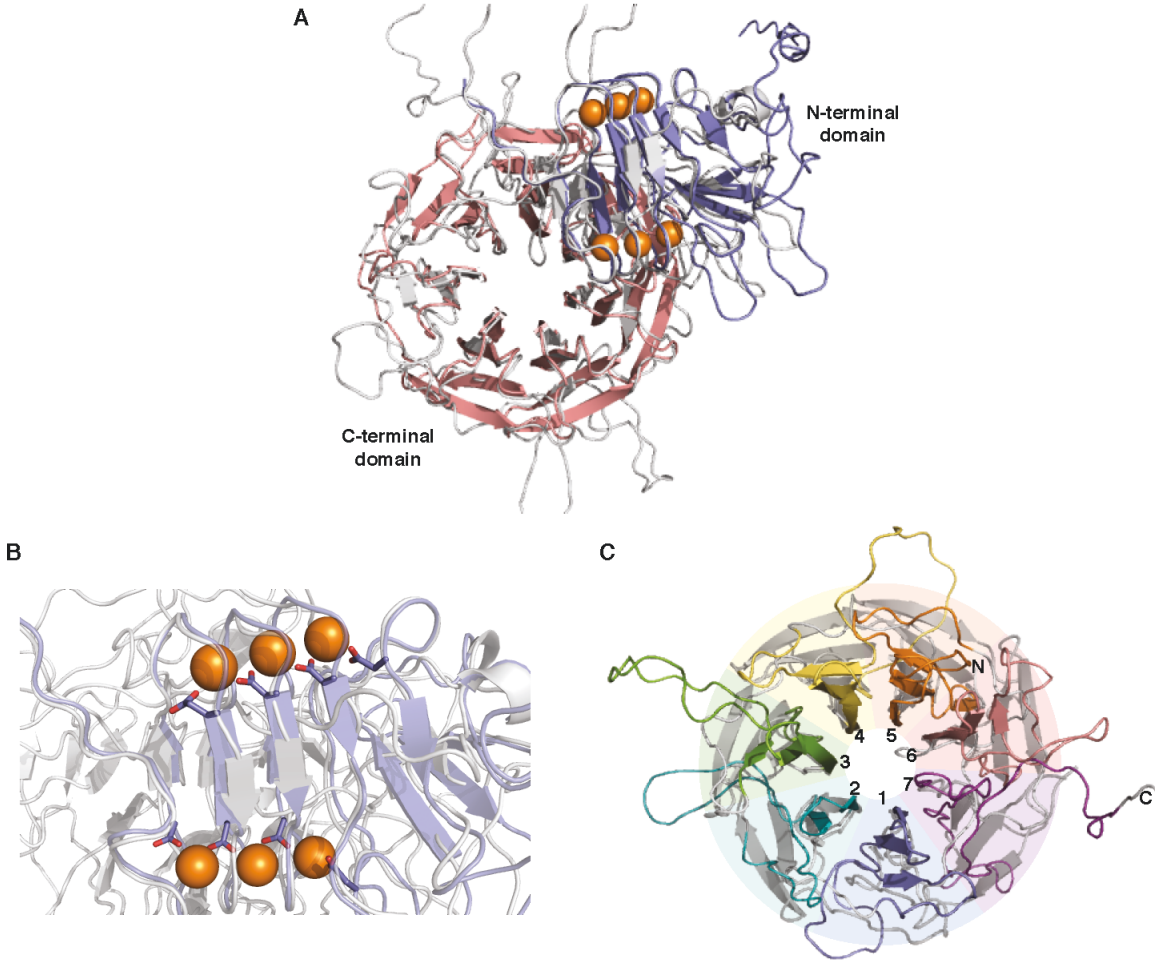
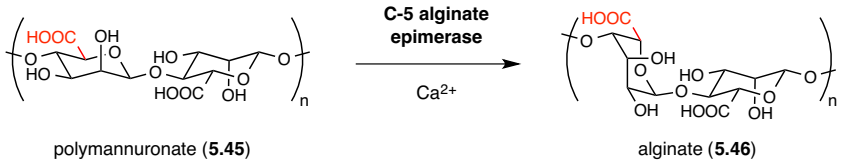
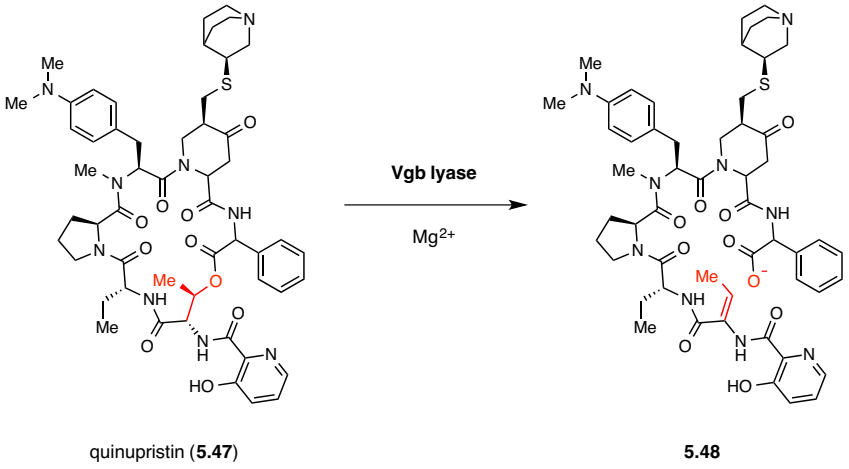


Figure 5.10: Homology model of CylK. **A**) Overall structure of CylK homology model (white) overlaid with the AlgE6 R-domain (blue) and Vgb lyase (pink). **B**) N-terminal calcium-binding domain of CylK (white) with the calcium-binding residues of the AlgE6 R-domain shown (blue). **C**) The C-terminal 7-bladed β-propeller NHL repeat domain of CylK overlaid with the Vgb lyase (grey).



Scheme 5.6: C-5 alginate epimerase catalyzes epimerization of polymannuronate to form alginates.



Scheme 5.7: The C–O bond cleavage of quinupristin catalyzed by the Vgb lyase.

The NHL repeats, each of which fold into a β-propeller blade, in CylK were identified in CylK based on the NCBI conserved domain analysis[27] and the alignment of the CylK homology model with the crystal structure of Vgb lyase.[36] Several characteristic sequence motifs are observed in the multiple sequence alignments of the seven NHL repeats/β-propeller blade regions in CylK (**Figure 5.11**). The NHL repeats (46-49 residue sequence) in CylK have a conserved GXTXG motif as well as several Asp/Glu, Gly and Lys residues that appear to be relatively well conserved near the C-terminal end. Each of the NHL repeats ends with a charged residue (Arg, Lys, Asp or Asn) and also contains an aromatic Phe/Trp residue upstream.

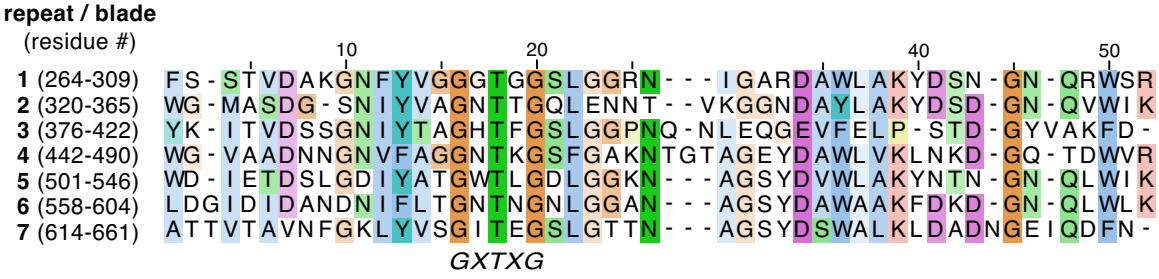


Figure 5.11: Multiple sequence alignment of the seven NHL repeats/blade regions of CylK.

5.2.3. CylK catalyzes the dimerization of chlorinated resorcinol to form cylindrocyclophane F

The secondary structural homologs of CylK, C-5 alginate epimerases[35] and Vgb lyase,[36] both catalyze enolate chemistry. The final dimerization step in the cylindrocyclophane biosynthesis is also expected to involve a phenolate chemistry with the resorcinol moiety acting as a nucleophile. CylK, therefore, is a good candidate enzyme for the C–C bond formation in the dimerization step. To study the function of CylK, the gene encoding CylK was cloned into *E. coli* for heterologous expression and purification. After screening various affinity tags, expression and purification conditions (**Table 5.2**), we successfully purified soluble CylK as a C-Strep-tag fusion protein (**Figure 5.12**).

Table 5.2: Conditions screened for the expression and the purifications of CylK
(Condition used for successful purification is bolded)

Affinity tag	N-His_6, C-His_6, N-Strep, **C-Strep**
Expression *E. coli* strain	BL21, **Rosetta**, Arctic Express
Expression temperature	**15 ºC**, 25 ºC, 37 ºC
Expression time	6-**21** h
Growth media	**LB with** or without **5-10 mM $CaCl_2$**
Lysis buffer	**HEPES** or Tris at pH 8.0 with or without $CaCl_2$

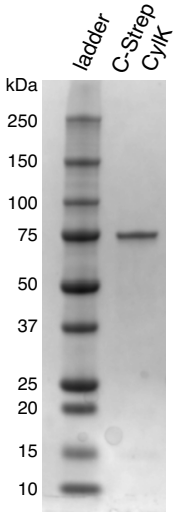
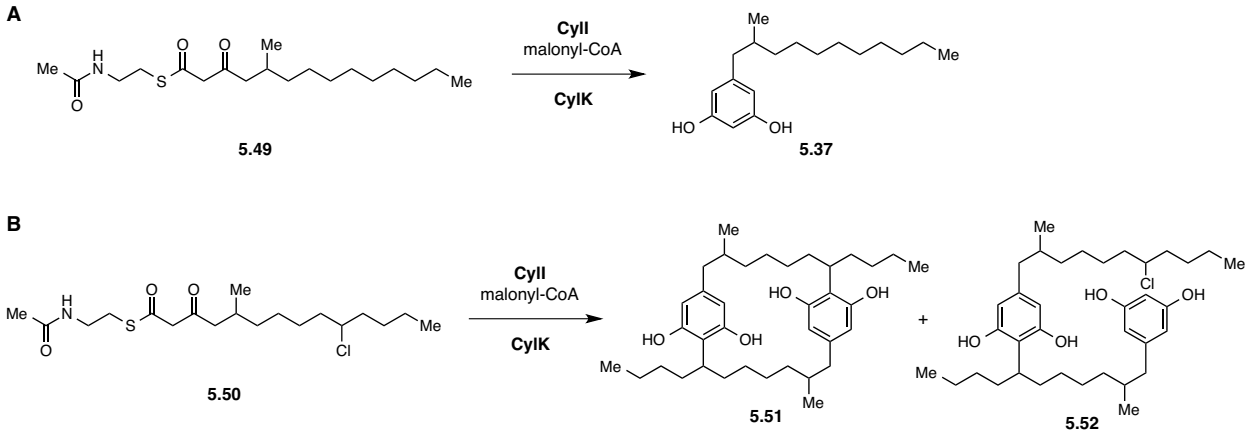


Figure 5.12: SDS-PAGE of purified C-Strep tagged CylK.

The activity of the purified CylK was tested in coupled assays with CylI using the synthetic β-ketoacyl-SNAC substrates (**5.49** and **5.50**, see **Scheme 5.8**). HPLC analysis of the assay samples revealed that the incubation of the chlorinated β-ketoacyl-SNAC substrate (**5.50**) in the presence of both CylI and CylK resulted in the formation of cylindrocyclophane F with undefined stereochemistry (**5.51**, see **Scheme 5.8B** and **Figure 5.13A**). Another major peak that formed in the full assay sample is predicted to be an intermediate (**5.52**) resulting from the formation of one C–C bond. This prediction was supported through detection of the mass corresponding to **5.52** in the assay samples using LC-HRMS (**Figure 5.13B**) and the fragmentation pattern observed in the MS/MS analysis (**Figure 5.13C**). Both cylindrocyclophane F (**5.51**) and **5.52** were absent in the negative controls, and the chlorinated resorcinol (**5.53**) accumulated in the assay lacking CylK. No CylK activity was observed in the assays performed with the non-chlorinated β-ketoacyl-SNAC substrate (**5.49**, see **Scheme 5.8A**).



Scheme 5.8: Coupled CylI and CylK assay using β-ketoacyl-SNAC substrates. **A**) Assay performed with the non-chlorinated substrate. **B**) Assay performed with the chlorinated substrate.

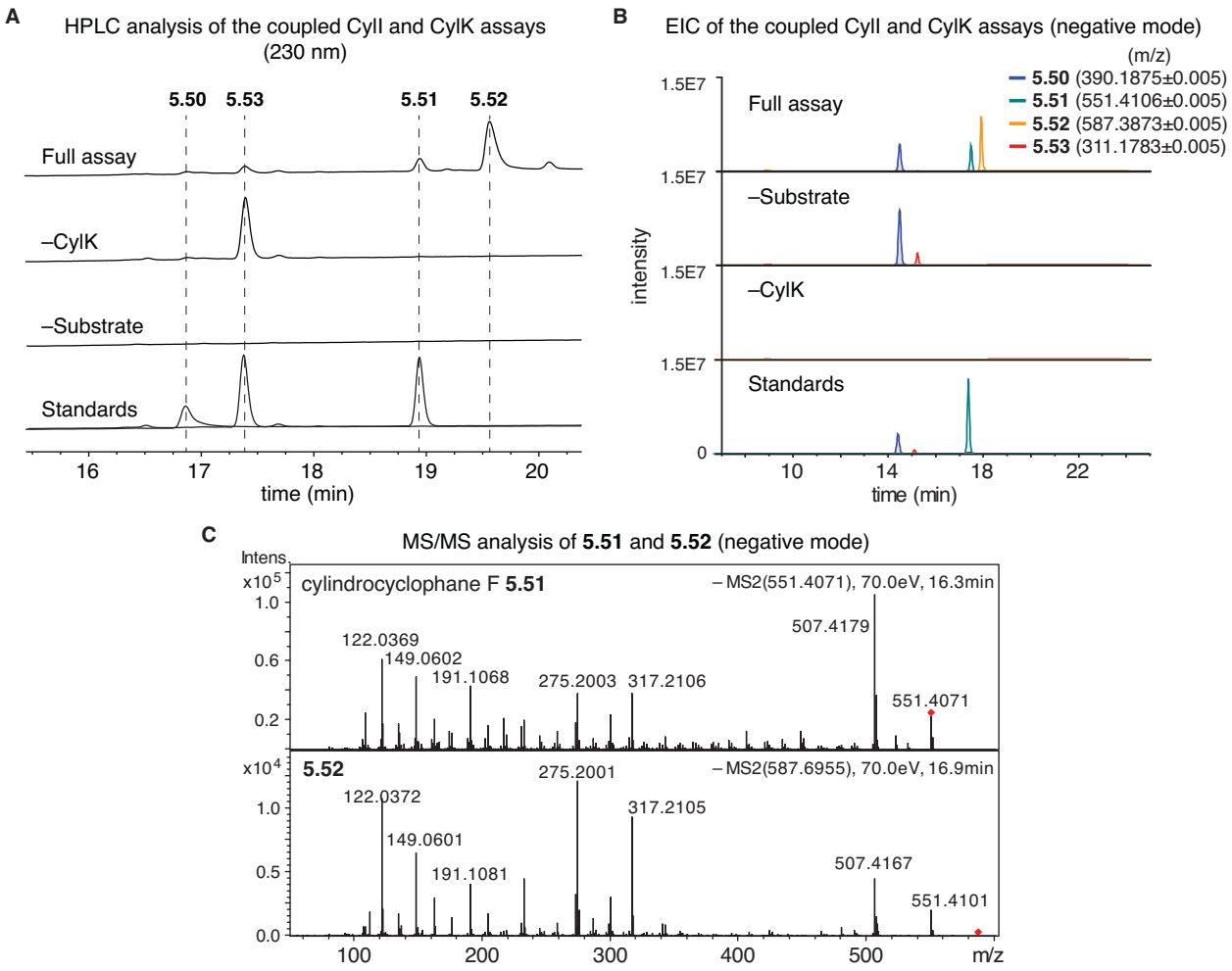
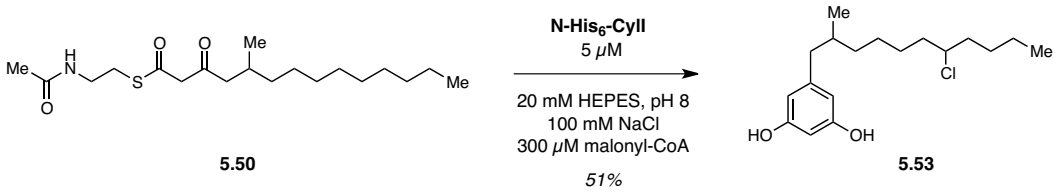


Figure 5.13: HPLC and LC-HRMS analysis of the coupled CylI and CylK assays. **A**) HPLC traces of the coupled assays monitored at 230 nm. **B**) LC-HRMS extracted ion chromatograms (EICs) of the coupled assays. **C**) Fragmentation patterns of cylindrocyclophane F (**5.51**) and **5.52** in the MS/MS analysis supports the predicted structure of **5.52**.

While we predict that CylK uses the chlorinated resorcinol to form cylindrocyclophane F, we cannot be certain what the exact substrate for the C–C bond formation is in the coupled assay. To determine whether CylK uses the chlorinated resorcinol (**5.53**) as a substrate, we performed large-scale CylI assays to generate and purify a diastereomeric mixture of the chlorinated resorcinol (**5.53**, see **Scheme 5.9**). The structure of the chlorinated resorcinol (**5.53**) was verified through ^{1}H, ^{13}C, COSY and HSQC NMR experiments (**Table 5.3** and **Figures 5.14-5.16**). Incubation of the purified chlorinated resorcinol (**5.53**) with CylK resulted in the formation of cylindrocyclophane F (**5.21**) and **5.52** (**Figure 5.17**). The assay result showed that CylK is capable of accepting the fully aromatized resorcinol as substrate to generate

one or two C–C bond(s). The use of the purified diasteremeric mixture of the chlorinated resorcinol (**5.53**) as substrate, however, leads mostly to accumulation of the intermediate (**5.52**) and only small amount of cylindrocyclophane F (**5.21**) formation is observed.



Scheme 5.9: Preparation of the chlorinated resorcinol (**5.53**) through large-scale CylI assays.

Table 5.3: NMR data of chlorinated resorcinol (**5.53**) in d_4-methanol (600 MHz).

Carbon	δ **C** (from HSQC)	δ **H, multiplicity** (*J* in Hz)	**COSY** (δ H)
1	not determined	-	-
2	102.3	6.08, s	-
3	not determined	-	-
4	109.9	6.10, s	-
5	not determined	-	-
6	109.9	6.10, s	-
7	46.0	2.45, dd (13.21, 8.00) 2.21, dd (13.22, 6.45)	2.21, 1.68 2.45, 1.68
8	37.2	1.68, m	2.21, 2.45, 0.85
9	21.2	0.85, d (6.61)	1.68
10	38.8	1.36, m	1.68, 1.28, 1.14
		1.14, m	1.58, 1.36, 1.28
11	32.0	1.28, m	1.50, 1.37, 1.36, 1.14
12,16	31.1, 29.0	1.50, m	1.72, 1.65, 1.37, 1.32, 1.28
		1.37, m	1.72, 1.65, 1.50, 1.32, 1.28
13,15	40.8 (2)	1.72, m	3.88, 1.65, 1.50, 1.37
		1.65, m	3.88, 1.72, 1.50, 1.37
14	66.3	3.88, m	1.72, 1.65
17	24.6	1.32, m	1.50, 1.37, 0.92
18	15.6	0.92, t (7.20)	1.32

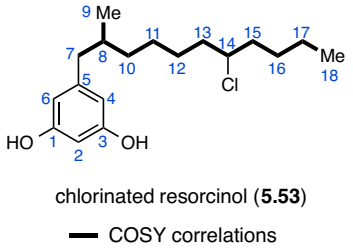


Figure 5.14: NMR characterizations of the chlorinated resorcinol (**5.53**). Bold bonds indicate COSY correlations.

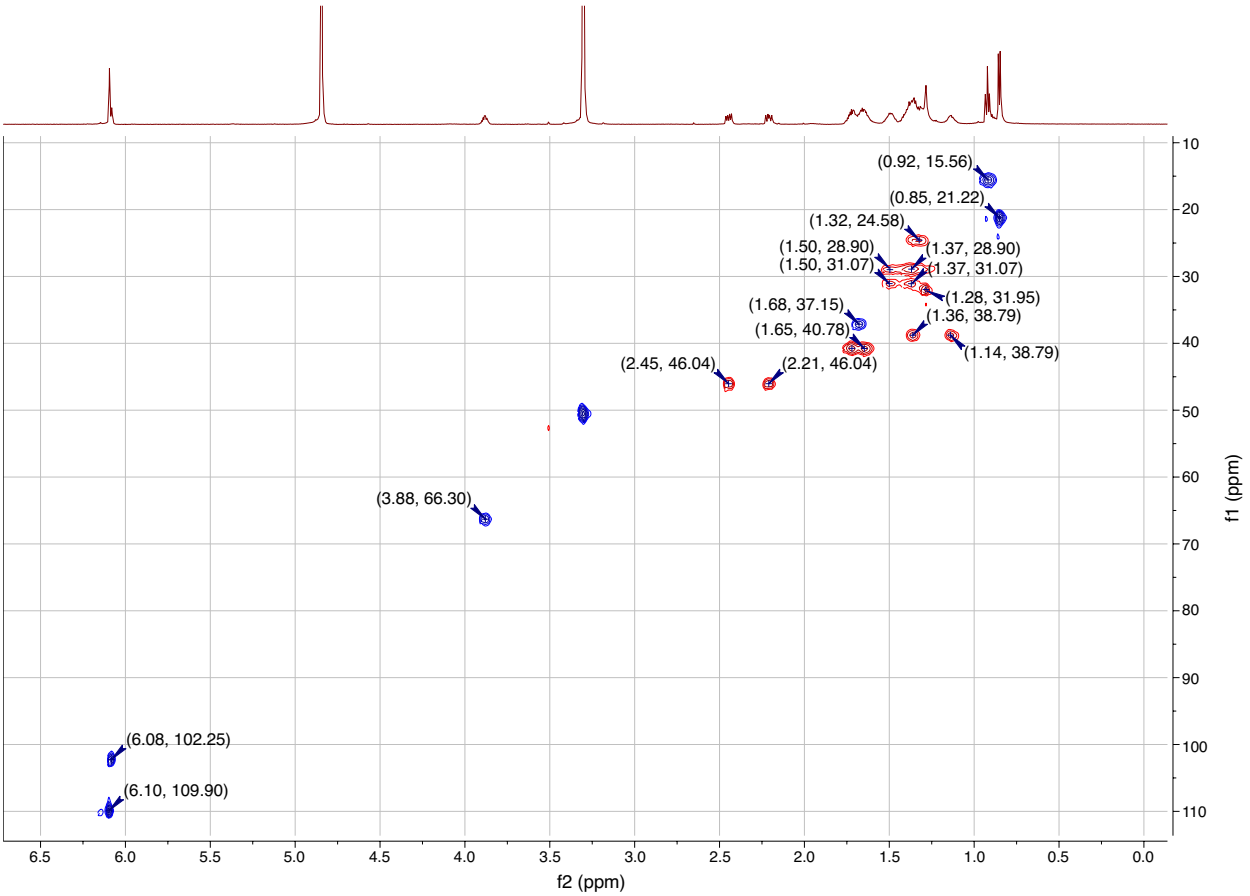


Figure 5.15: HSQC spectrum of the chlorinated resorcinol (**5.53**) in d_4-methanol (600 MHz). Blue signals are attributed to CH or CH_3 groups and red signals to CH_2 groups.

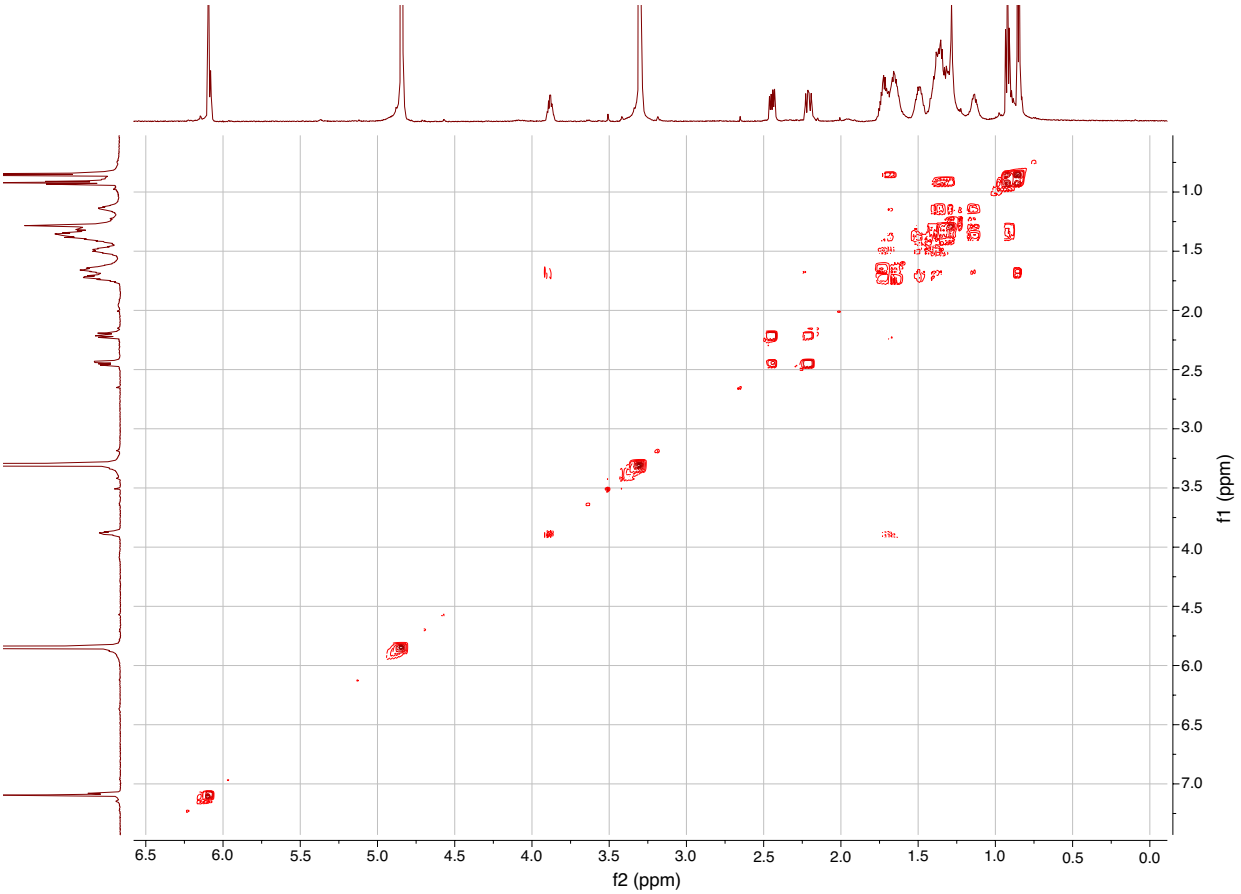


Figure 5.16: COSY spectrum of the chlorinated resorcinol (**5.53**) in d_4-methanol (600 MHz).

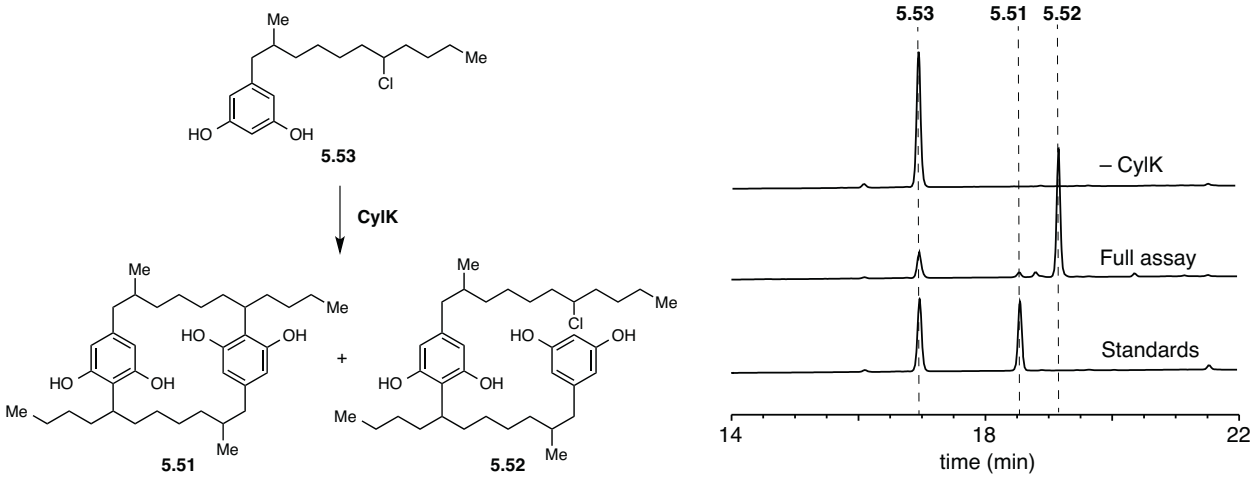


Figure 5.17: HPLC analysis of the CylK assay using the chlorinated resorcinol (**5.53**) as a substrate (monitored at 210 nm).

5.2.4. Determining the metal dependency of CylK

The bioinformatic analyses indicate that CylK is likely a calcium-dependent enzyme. To determine whether CylK required Ca^{2+} or other divalent metals for catalytic activity, we first compared the activity of CylK in buffers with and without 10 mM $CaCl_2$ and $MgCl_2$. The CylK activity observed in the coupled CylI and CylK assays is comparable with and without excess $CaCl_2$ and $MgCl_2$. Next we tested whether the activity of CylK changes in the presence of EDTA, which sequesters metal cations. The addition of EDTA to the coupled CylI and CylK assay resulted in the loss of CylK activity, which indicates that metal ions are required for the C–C bond formation (**Figure 5.18**).

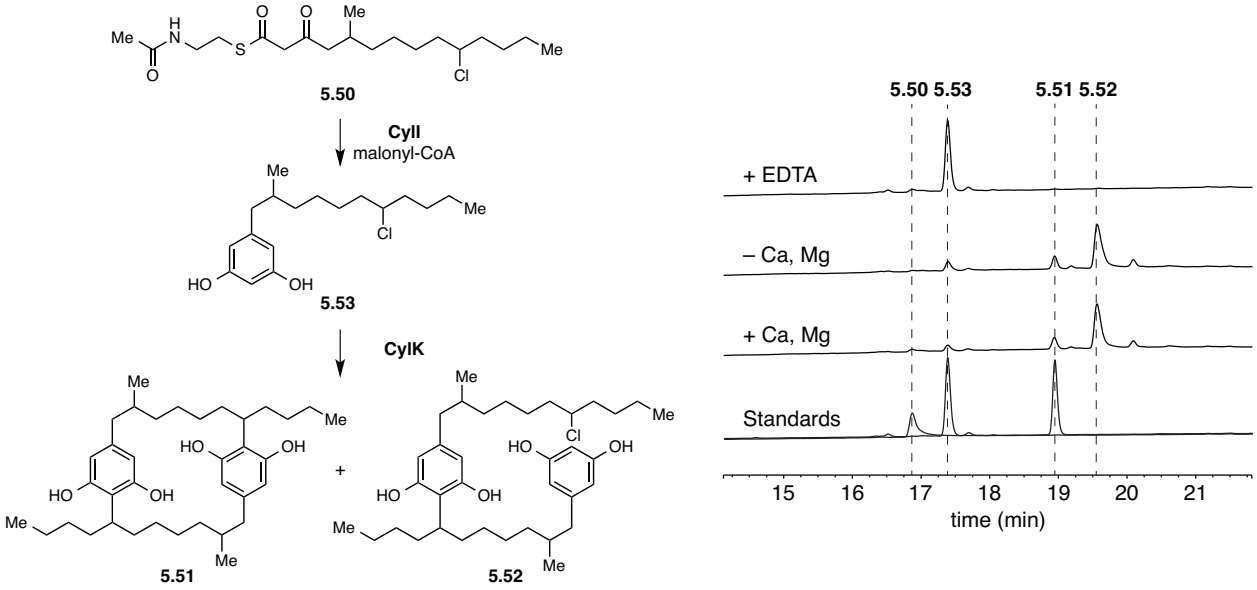


Figure 5.18: Determining the metal-dependency of the CylK activity in coupled CylI and CylK HPLC assays (monitored at 230 nm).

To identify the divalent metal ions required for the CylK activity, an aliquot of the purified CylK was incubated with 2 mM EDTA and purified by gel filtration FPLC into a buffer without $CaCl_2$ and $MgCl_2$. The gel filtration FPLC analysis revealed that the removal of metals using EDTA resulted in a shift in the oligomeric state from monomer to mostly dimer (**Figure 5.19**). The shift in the oligomeric state was not observed when the untreated CylK was purified by gel filtration FPLC using the buffer without $CaCl_2$ and $MgCl_2$. The result of the coupled assays and the gel filtration FPLC analysis indicate that the metal ions

required for the CylK activity are bound tightly, with EDTA-treatment required to remove the metal ions from the enzyme.

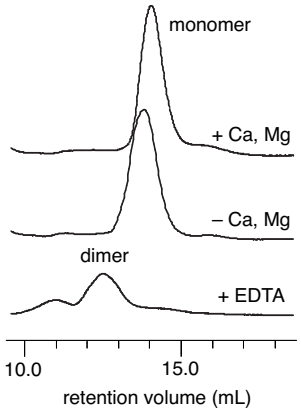


Figure 5.19: The analysis of the CylK oligomeric state using gel filtration FPLC.

Using CylK treated with EDTA and purified by FPLC, we screened the dependency of the CylK activity on various divalent metals (Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+} and Zn^{2+}) by adding 10 mM of each metal to the assay mixture. The result of the screen showed that only Ca^{2+} is capable of restoring the activity of the EDTA-treated CylK (**Figure 5.20A**). In addition, we tested the activity of CylK in presence of other divalent metal ions in combination with Ca^{2+} (**Figure 5.20B**). The analysis shows that the addition of Zn^{2+} (and Fe^{2+} to a lesser extent) inhibits the activity of CylK, possibly due to competitive binding to the metal-binding sites. While other divalent metal ions (Mg^{2+}, Mn^{2+}, Fe^{2+} and Zn^{2+}) do not appear to be required for the CylK activity as long as sufficient Ca^{2+} is provided, we cannot exclude the possibility that the EDTA-treatment does not remove all the metal ions that may influence the enzyme activity. To determine which divalent metal ions are bound to CylK, we performed ICP-MS analysis on the sample of CylK dialyzed against the buffer lacking divalent metals. The analysis indicated that only Ca^{2+} (2.9 equiv.) and Mg^{2+} (1.6 equiv.) are present in the CylK sample, and the other divalent metals (Fe^{2+}, Mn^{2+} and Zn^{2+}) are absent. The metal dependency assays and the ICP analysis confirmed that CylK is a calcium-dependent enzyme as predicted by the bioinformatic analyses. Although the role of Ca^{2+} is yet to be determined, we

predict that Ca^{2+} is either bound to the active site for catalysis or allosterically bound to regulate the activity of CylK. Alternatively, Ca^{2+}-binding might only be required for the structural integrity of CylK.

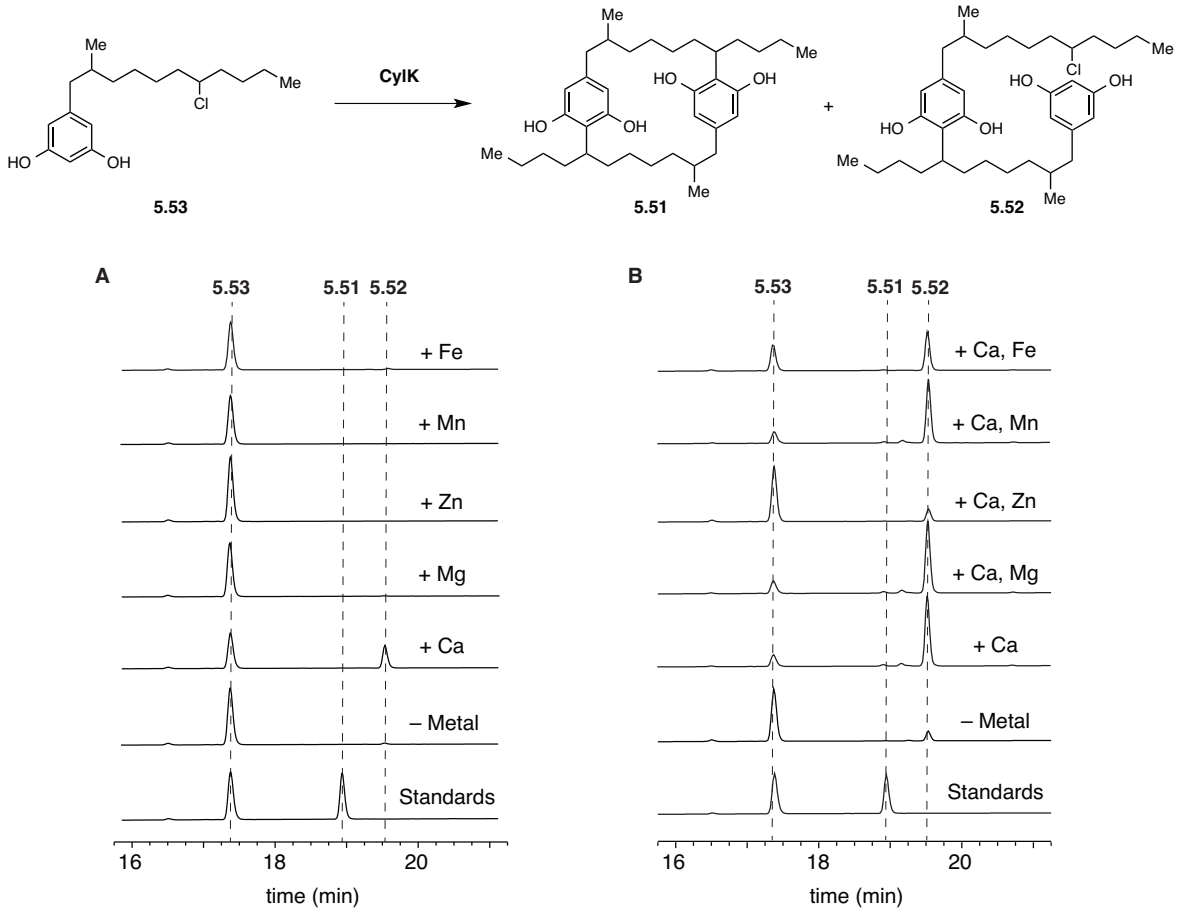
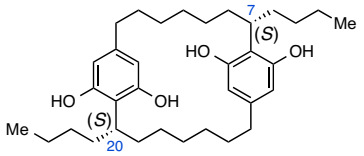


Figure 5.20: Screening the metal dependency of CylK by HPLC assays (monitored at 210 nm). **A**) Addition of various divalent metal ions to the assay. **B**) Testing the combination of various divalent metal ions with Ca^{2+}.

5.2.5. Stereochemical analysis of the CylK catalyzed C–C bond formation

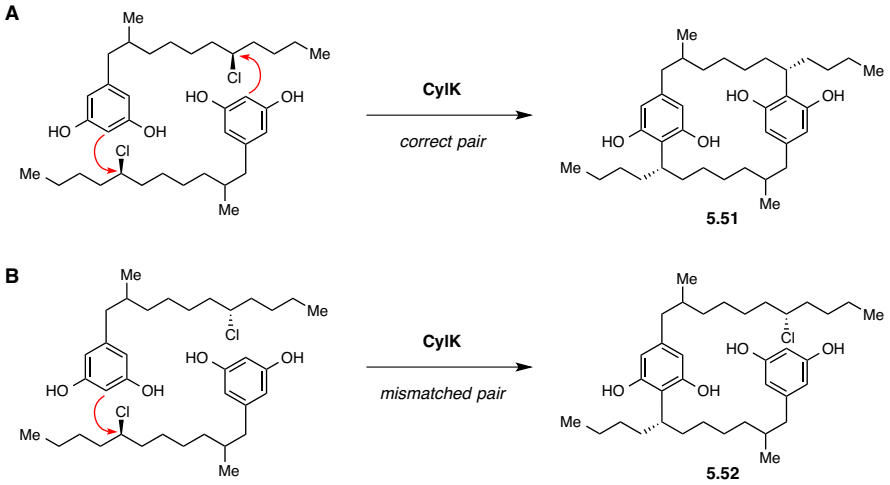
One important question about the CylK catalyzed C–C bond formation is whether this reaction is stereoselective or not. All the isolated natural products with the [7.7]paracyclophane scaffold appear to have absolute *S* configuration at C7/C20, which are the carbon centers involved in the final dimerization step (**Figure 5.21**).[13,16-18,43] The strict conservation of the stereochemical confirugration at C7/C20 suggests that either the dimerization step or the chlorination step (or both) is stereoselective.



core [7.7]paracyclophane scaffold
(**5.54**)

Figure 5.21: The isolated natural products with the [7.7]paracyclophane scaffold have a conserved stereochemical configuration.

One observation that supports the proposal that CylK catalyzes stereoselective (or stereospecific) C–C bond formation is the accumulation of **5.52** in the CylK assays performed with the diastereomeric mixture of the chlorinated resorcinols (**5.53**). If CylK selectively forms C–C bonds using alkyl chlorides using alkyl chlorides of only one stereochemical configuration, the intermediate **5.52** may accumulate in the assay mixture when monomers with the wrong stereochemistry are coupled; this 'mismatched' pair would be unable to undergo a second alkylation (**Scheme 5.10**). The accumulation of **5.52** as well as the presence of the other diastereomers may inhibit CylK, leading to the low conversion to the final macrocyclic product (**5.51**).



Scheme 5.10: Possible outcomes of combining different diastereomers of the chlorinated resorcinol. **A**) Coupling of two units of the correct diastereomers. **B**) Coupling of the correct and the incorrect diastereomers.

To determine whether CylK selectively uses one diastereomer over the others, we performed a chiral HPLC analysis of the chlorinated resorcinol (**5.53**) that remained in the non-coupled CylK assay mixture

after no further conversion was observed (**Figure 5.22**). The chlorinated resorcinol substrate (**5.53**), which was purified from the CylI assays, contained all four possible diastereomers (**5.55**-**5.58**), indicating that the CylI does not have a strong preference for any particular substrate steroisomers. After the CylK assay, however, only two out of the four resorcinol diastereomers remained. This observation indicates that CylK selectively uses two diastereomers and the C–C bond formation catalyzed by CylK is stereoselective. We hypothesize that CylK is capable of using the two diastereomers with the correct stereochemical configuration at the chlorinated carbon center. The stereochemistry of the methyl substituent is likely not as important for CylK selectivity, since that stereocenter is relatively remote from the actual site of catalysis.

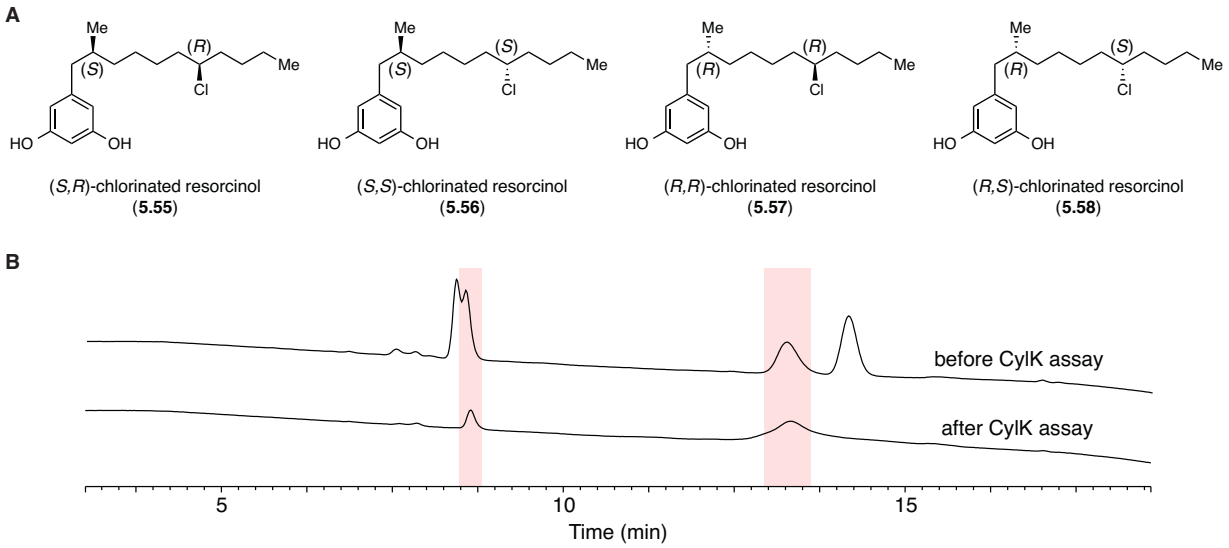


Figure 5.22: Stereochemical analysis of the chlorinated resorcinol. **A**) Structures of the four diastereomers. **B**) Chiral HPLC analysis of the chlorinated resorcinol before and after the CylK assay (monitored at 230 nm).

To determine which diastereomers CylK preferentially uses, we decided to synthesize diastereomerically-enriched substrates. Due to synthetic feasibility, we first synthesized two diastereomers of the chlorinated acyl-SNAC (**5.59** and **5.60**, see **Figure 5.23**), which serve as substrates for the type I PKS CylH. Based on this assumption, we synthesized the two diastereomers with the *S* configuration at the methylated carbon center, which is the absolute configuration observed in the naturally occurring

cylindrocyclophanes (**5.14**-**5.19**). Chiral HPLC analysis confirmed that each of the acyl-SNAC substrates (**5.59** and **5.60**) is primarily a single diastereomer with >96% diastereomeric eccess (**Figure 5.24**).

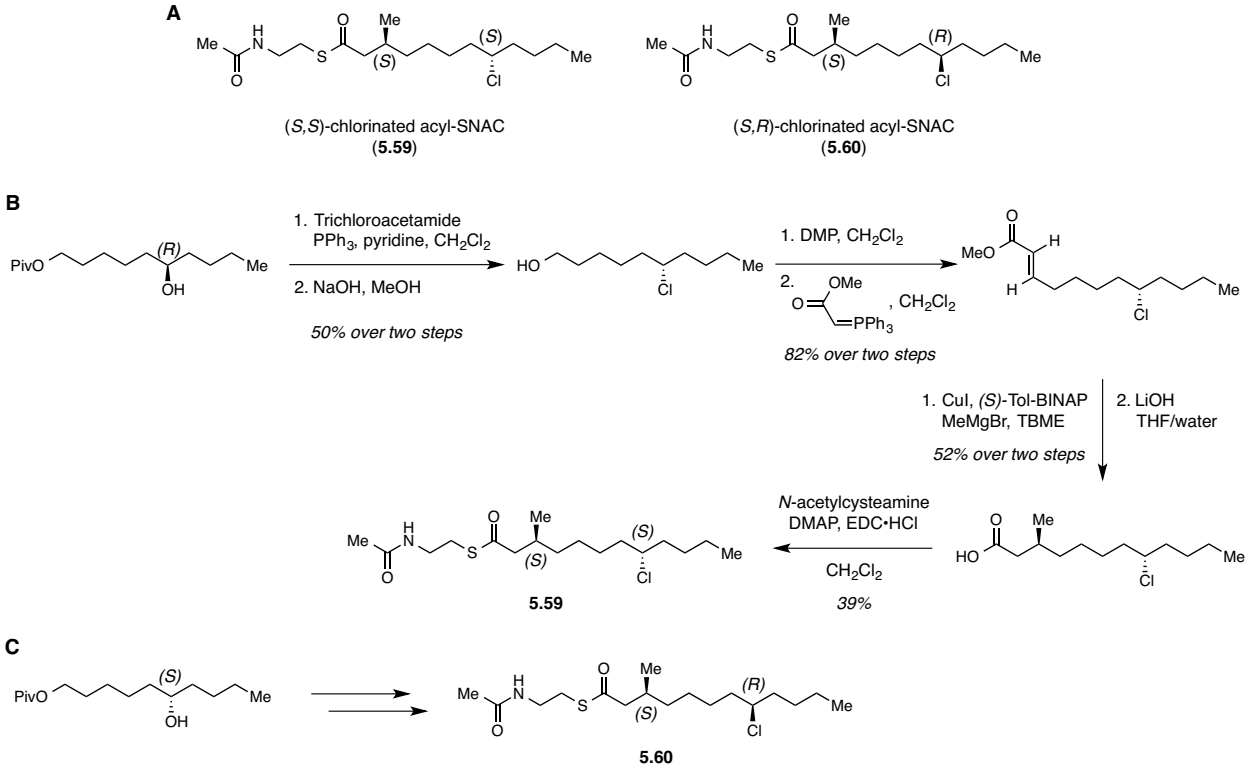


Figure 5.23: Synthesis of the diastereomerically enriched chlorinated acyl-SNAC substrates for CylH. **A**) Structures of the two diastereomers synthesized. **B**) Synthetic route for the (*S,S*)-chlorinated acyl-SNAC substrate **5.59**. **C**) The (*S,R*)-chlorinated acyl-SNAC substrate **5.60** was synthesized following the same route as the (*S,S*) diastereomer starting the other enantiomer.

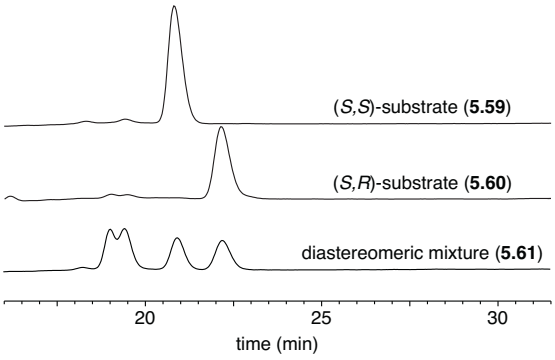
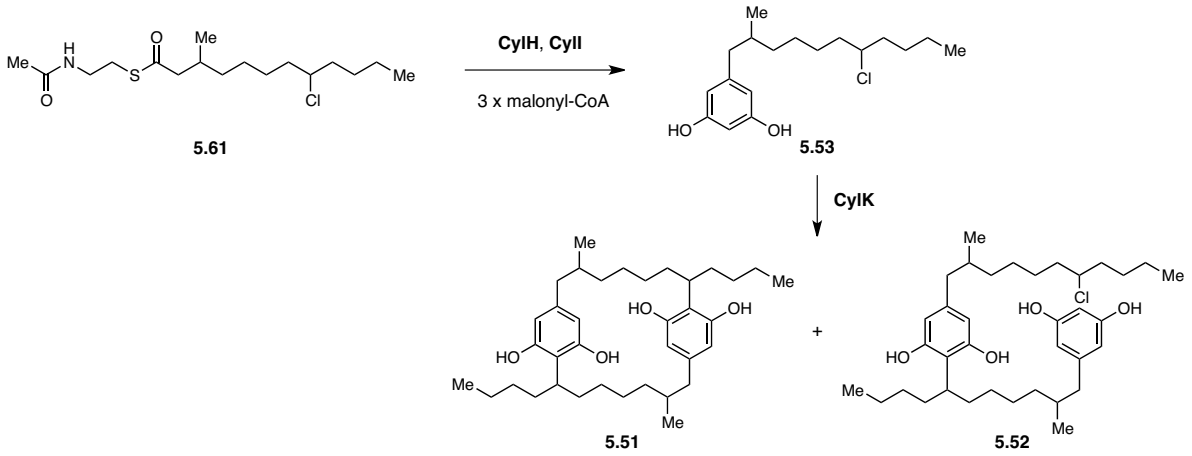


Figure 5.24: Reverse phase chiral HPLC analysis of the synthetic acyl-SNAC substrates (monitored at 230 nm).

We used the stereodefined, diastereomerically-enriched acyl-SNAC substrates (**5.59** and **5.60**), as well as the mixture of the four diastereomers (**5.61**), in coupled CylH, CylI and CylK assays (**Scheme 5.11**). The use of the diastereomeric mixture (**5.61**) in the three-enzyme coupled assays resulted in the accumulation of the intermediate **5.52** as was observed in the assays performed with the diastereomeric mixture of the chlorinated resorcinol (**Figure 5.25**). The use of the (*S,S*)-substrate (**5.59**) resulted in the accumulation of the (*S,S*)-chlorinated resorcinol (**5.56**), which indicates that CylH and CylI can use acyl-SNAC substrate **5.59** to form the chlorinated resorcinol (**5.56**), but CylK is unable to use the (*S,S*)-chlorinated resorcinol (**5.56**) as a substrate (**Scheme 5.12A**). On the other hand, the use of the (*S,R*)-substrate (**5.60**) resulted in the formation of cylindrocyclophane F (**5.19**) as the major product, and only small amount of the intermediate **5.52** was observed (**Scheme 5.12B**). The high stereoselectivity of CylK observed in the three-enzyme coupled assays might indicate that the CylK catalyzed C–C bond formation is stereospecific.



Scheme 5.11: Coupled CylH, CylI and CylK assay (shown for the mixture of diastereomers **5.61** as substrate).

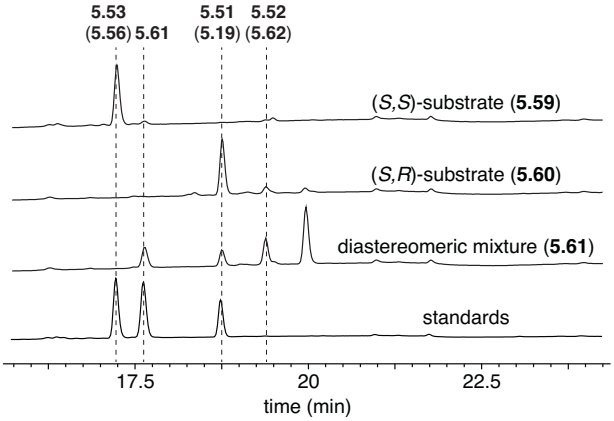
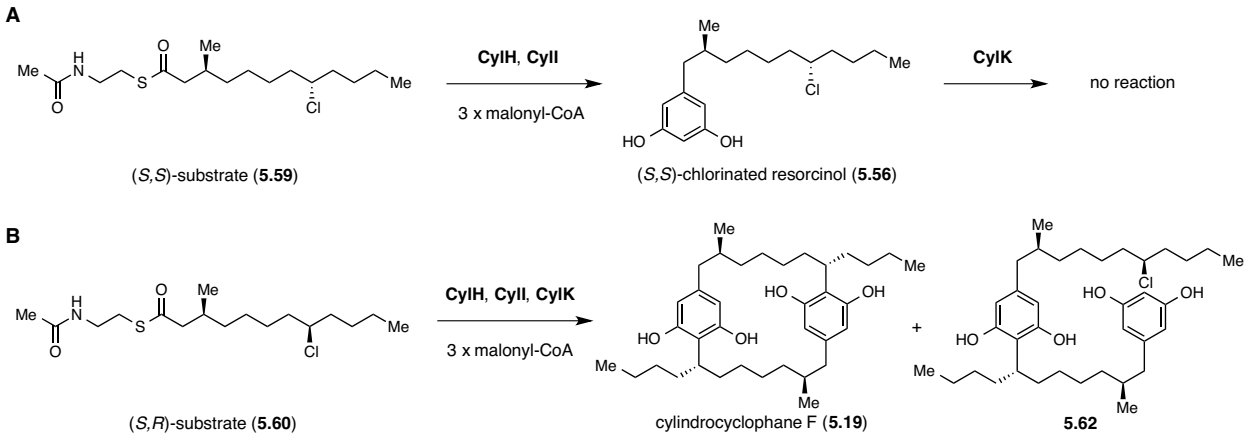


Figure 5.25: HPLC analysis of the coupled CylH, CylI and CylK assays (monitored at 230 nm).

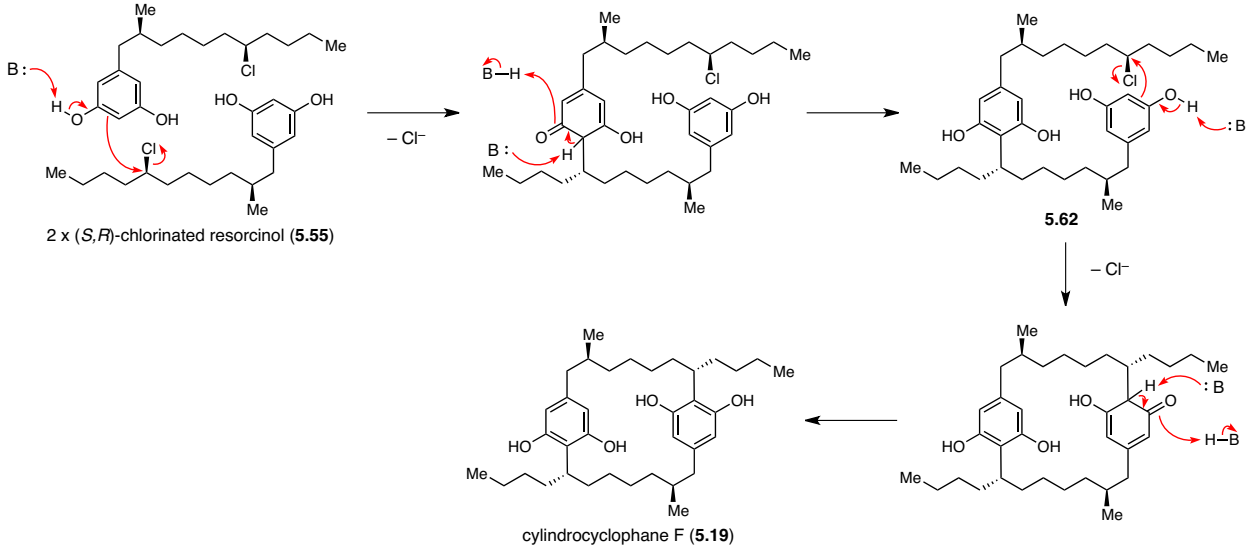


Scheme 5.12: Outcomes of the coupled CylH, CylI and CylK assays using diastereomerically-enriched substrates. **A**) Assay performed with the (*S,S*)-substrate. **B**) Assay performed with the (*S,R*)-substrate.

Based on the chiral HPLC analysis of the CylK assays and the result of the coupled assays using the stereodefined substrates, CylK is capable of using the chlorinated resorcinol with the *R* configuration at the chlorinated carbon center (**5.55** and **5.57**). The use of the substrate with the *R* stereochemistry shows that CylK-catalyzed C–C bond formation results in an inversion of configuration at C7/C20, based on the known absolute stereochemistry of the cylindrocyclophanes.[13,16-18,43] While CylK has a strict preference for the configuration of the chlorine substituent, it appears to have a relatively relaxed selectivity for the configuration of the methyl substituent. This relaxed selectivity is likely required to accommodate the resorcinol substrates with benzylic hydroxylation or acetylation for the production of cylindrocyclophanes A-E (**5.15**-**5.18**).

5.2.6. Predicted mechanism of the CylK catalyzed dimerization event

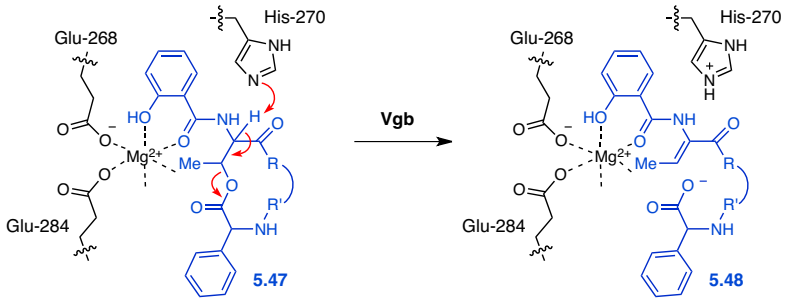
The preferred substrate stereochemistry of CylK indicates that this enzyme catalyzes the dimerization of the (*S,R*)-chlorinated resorcinol units (**5.55**) to form the [7.7]paracyclophane ring. This preference is consistent with alkylation occuring via an S_N2 mechanism, which would require inversion of stereochemistry with the leaving chloride group oriented *anti* to the C–C bond being formed. We hypothesize that CylK catalyzes the S_N2 reaction through acid-base chemistry using the phenolate of the resorcinol moiety as a nucleophile and the alkyl chloride as an electrophile (**Scheme 5.13**).



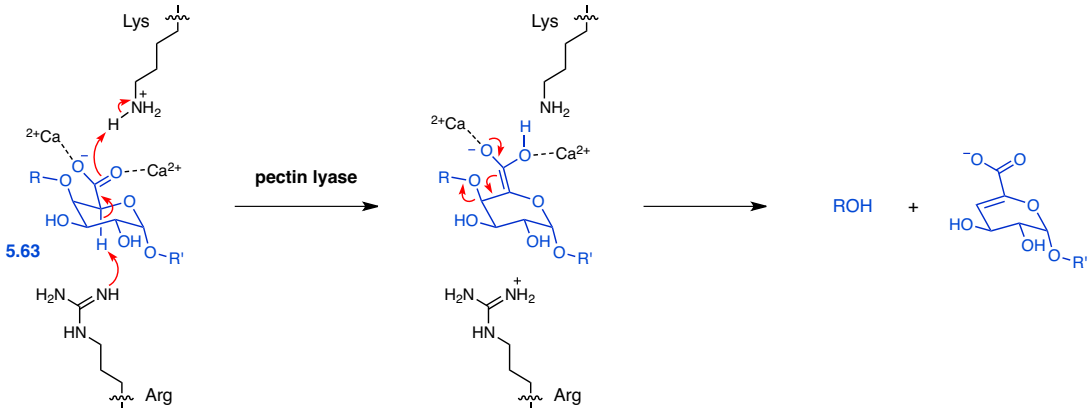
Scheme 5.13: Proposed mechanism of the CylK-catalyzed dimerization step.

CylK might use metal cations, possibly Mg^{2+} or Ca^{2+} based on the ICP-MS data, for catalysis. Examples of the enzymes that require Mg^{2+} or Ca^{2+} for catalysis are the Vgb lyase[36,42] and the pectin lyase.[44,45] The Vgb lyase requires Mg^{2+} for catalysis, and the crystal structure of this enzyme showed that the Mg^{2+} cation is coordinated to the oxygen atoms of the conjugated 3-hydroxypicolonic acid group of the substrate **5.47** (**Scheme 5.14**).[36] The Lewis acidity of Mg^{2+} is thought to be important for lowering the pK_a of the □-proton, which is deprotonated by the basic His-270 residue.[36] The pectin lyases PelC[44] and BsPel[45] have been shown to bind multiple Ca^{2+} ions for coordinating the negatively charged polygalacturonate substrate (**5.63**) in their active sites (**Scheme 5.15**). The Ca^{2+} ions are thought to be required for lowering

the pK_a of the α-proton for deprotonation by the catalytic arginine residue.[44,45] In addition, Ca^{2+} is also suggested to be important for stabilizing the alkoxide leaving group to promote the *anti*-β-elimination.[45] As in the Vgb and the pectin lyase cases, CylK may require metal cations in the active site to coordinate the resorcinol enolate and/or to stabilize the chloride leaving group. Moving forward, obtaining the structural information on the enzyme will be important for determining the role of the metal cations and for elucidating the mechanism of the CylK-catalyzed C–C bond formation.



Scheme 5.14: Vgb-catalyzed C–O bond cleavage requires Mg^{2+} for activity.

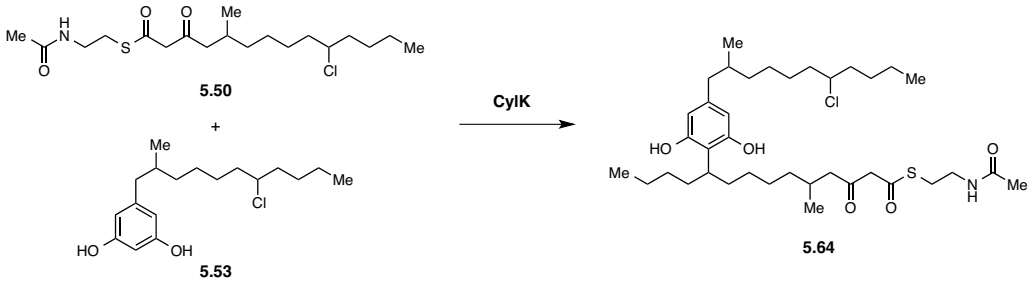


Scheme 5.15: Pectin lyase catalyzes *anti*-β-elimination of pectin (polygalacturonate) using Ca^{2+}.

5.2.7. Substrate scope analysis of the CylK catalyzed C–C bond formation

Another property of CylK we wished to screen was the substrate scope. The analysis of the initial CylI and CylK coupled assay by LC-HRMS revealed that CylK is also capable of forming a C–C bond between the chlorinated β-ketoacyl-SNAC substrate **5.50** and the chlorinated resorcinol **5.53** (**Scheme**

5.16 and **Figure 5.26**). This observation suggested that CylK is relatively promiscuous and might have an expanded substrate scope.



Scheme 5.16: CylK catalyzes C–C bond formation between the chlorinated -ketoacyl-SNAC substrate (**5.50**) and the chlorinated resorcinol (**5.53**).

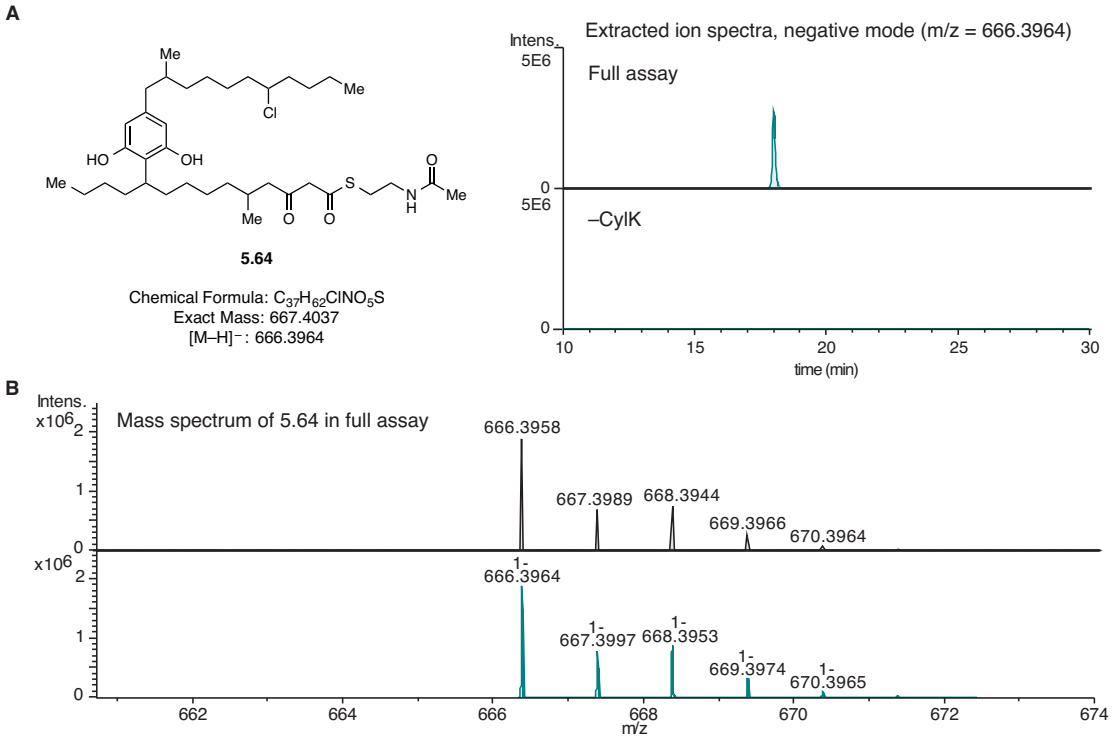
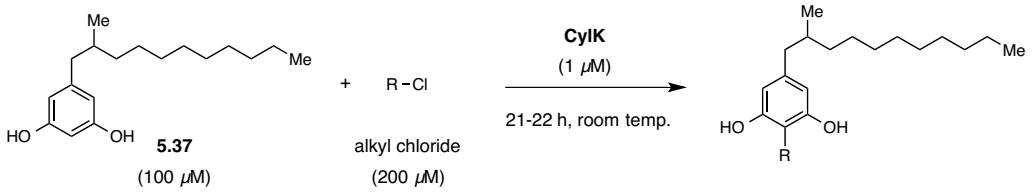


Figure 5.26: LC-HRMS detection of **5.64**. **A**) Extracted ion chromatograms of **5.64**. **B**) The mass spectra of **5.64**.

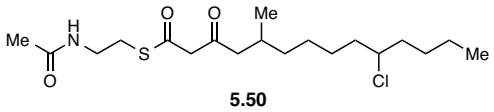
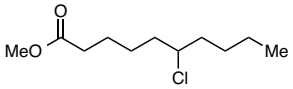
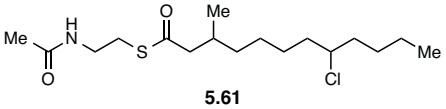
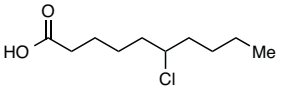
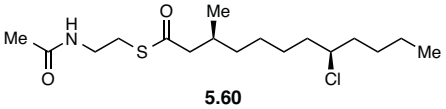
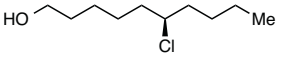
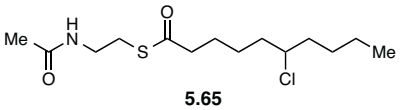
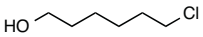
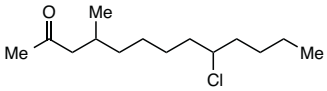
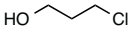
To analyze the substrate scope of CylK, we first screened the ability of CylK to catalyze C–C bond formation between the non-chlorinated resorcinol (**5.37**) and various alkyl chlorides (**Scheme 5.17**). Each alkyl chloride substrate (**5.50**, **5.60**, **5.61**, **5.65**-**5.71**) was incubated with resorcinol **5.47** and CylK for 21–22 hours at room temperature. The coupled products (**5.72**-**5.77**, see **Figure 5.27**) were then detected by

HPLC (**Figure 5.28**, **Table 5.4**). The reactions were performed with resorcinol **5.37** as the limiting agent. The conversion (%) was calculated from the peak areas of the product and the unreacted resorcinol **5.37** at wavelengths with minimal absorbance for the alkyl chloride substrates (210 or 280 nm).



Scheme 5.17: CyK-catalyzed alkylation of the non-chlorinated resorcinol **5.37** with various alkyl chloride substrates.

Table 5.4: HPLC conversions (%) of the CylK-catalyzed resorcinol alkylation with various alkyl chlorides.

alkyl chloride substrate	conversion	alkyl chloride substrate	conversion
 **5.50**	87 %	 **5.67**	14 %
 **5.61**	91 %	 **5.68**	0 %
 **5.60**	93 %	 **5.69**	51 %
 **5.65**	56 %	 **5.70**	0 %
 **5.66**	74 %	 **5.71**	0 %

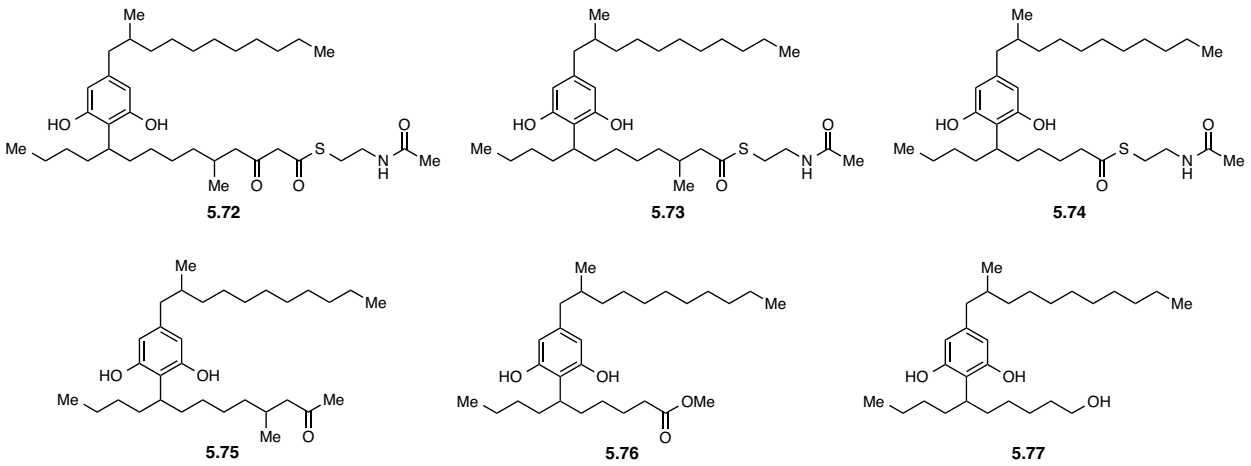


Figure 5.27: Predicted structures of the products of the CylK-catalyzed resorcinol alkylation assays in which product formation was observed.

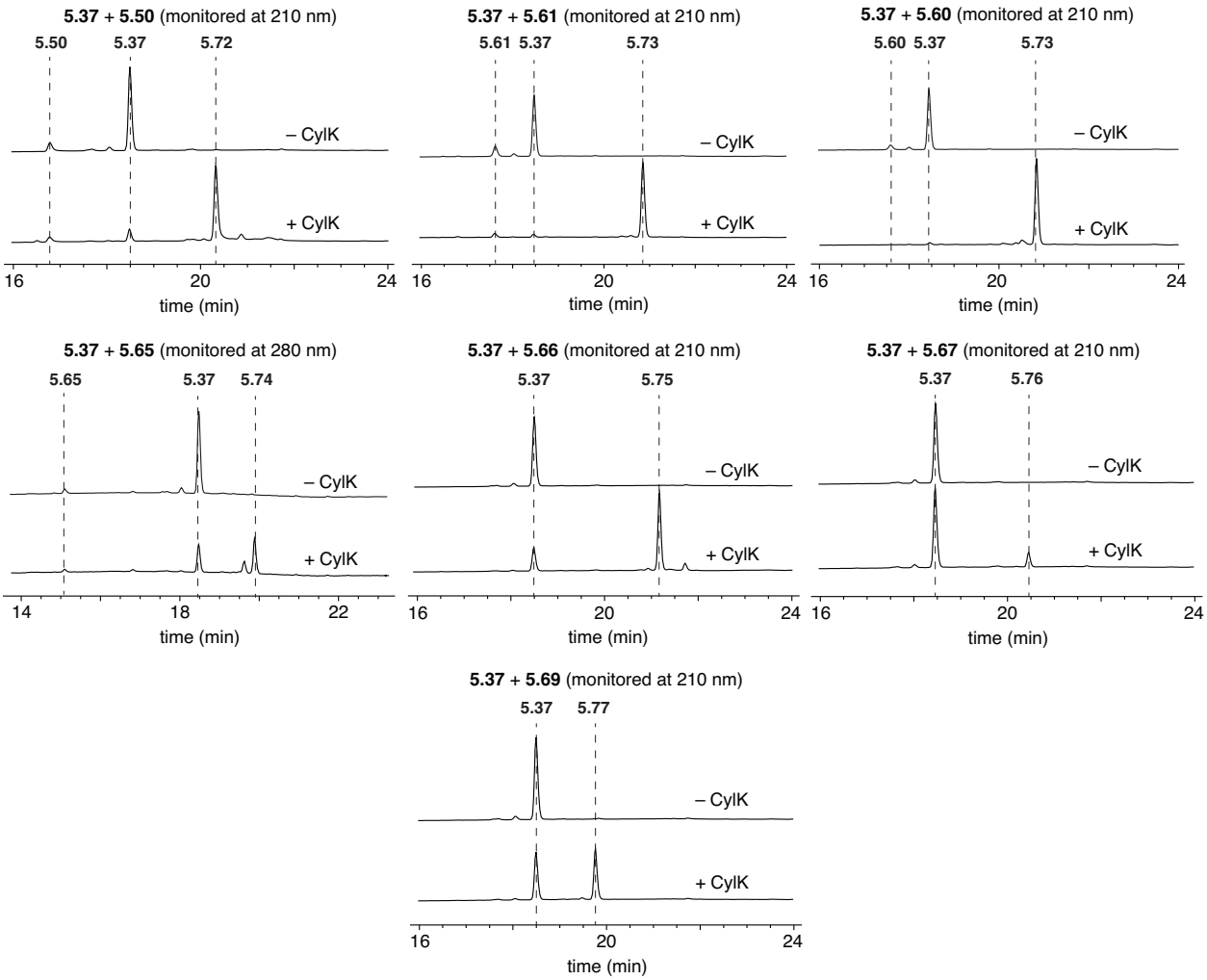
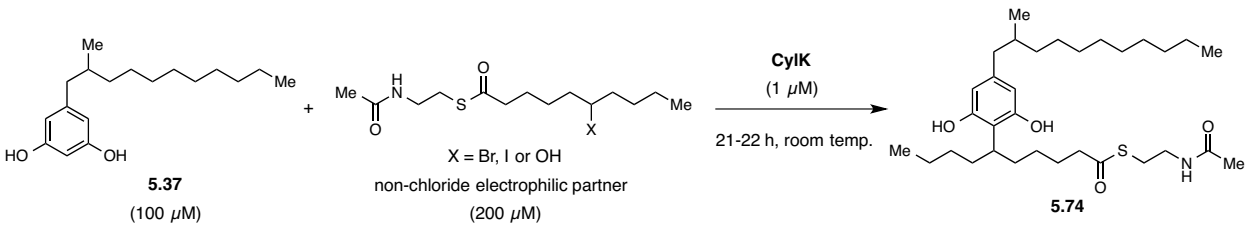


Figure 5.28: HPLC analysis of the CylK-catalyzed resorcinol alkylation assays with the product formation observed.

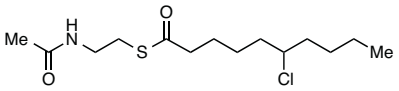
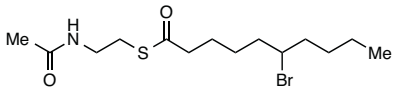
As we anticipated, CylK is capable of reacting the non-chlorinated resorcinol (**5.37**) with many of the alkyl chlorides we screened. All the SNAC substrates (**5.50**, **5.60**, **5.61** and **5.65**) we tested showed relatively high conversions (56-93%). While the product formation was also observed for other carbonyl substrates, 74% for the ketone (**5.66**) and 14% for the methyl ester (**5.67**), no conversion was observed for the carboxylic acid substrate (**5.68**). The lack of conversion with the carboxylic acid substrate (**5.68**) could be due to the negative charge on the molecule at pH 8, which likely leads to charge repulsion in the active site. Lastly, 51% conversion was observed for 6-chlorodecanol (**5.69**) but no reaction occurred using 1-chlorohexanol (**5.70**) or chloropropanol (**5.71**). This outcome suggests that either CylK is incapable of using primary alkyl chlorides as electrophiles or the butyl tail is important for substrate binding.

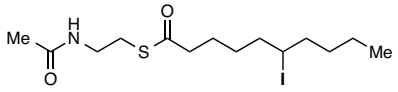
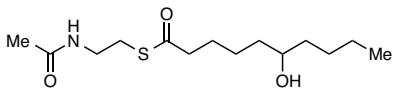
Another question we wanted to explore was whether CylK was capable of coupling resorcinol to other electrophiles containing different leaving groups. The CylK activity toward additional electrophiles was tested using the decanoyl-SNAC derivatives (**5.78**-**5.80**, see **Scheme 5.18**, **Figure 5.29** and **Table 5.5**). Interestingly, the conversion of the alkyl bromide substrate (**5.78**) was comparable to that of the alkyl chloride substrate (**5.65**), which agrees with the successful isolation of cylindrocyclophanes and its analogs from the cultures of native producers grown in bromide-enriched media.[16,17] Some conversion was also observed with the iodide substrate (**5.79**), which shows that CylK is promiscuous toward different halide leaving groups. CylK was unable to use the alcohol (**5.80**) as substrate, likely because hydroxide is a poor leaving group. The inability of CylK to catalyze a C–C bond formation using the alcohol (**5.80**) indicates that if there is a metal cation in the enzyme active site, as in the case of pectin lyase, the metal cation does not activate the hydroxyl group to act as a leaving group. The observed substrate preference confirms that alkyl chloride is the natural substrate of CylK, although alkyl bromide and alkyl iodide, to a lesser extent, can also serve as substrates.



Scheme 5.18: CylK assay with non-chloride electrophilic partners.

Table 5.5: HPLC conversion (%) of the CylK-catalyzed resorcinol alkylation with other alkyl halides or alcohols.

electrophilic partner	conversion
 **5.65**	56 %
 **5.78**	47 %

electrophilic partner	conversion
 **5.79**	7 %
 **5.80**	0 %

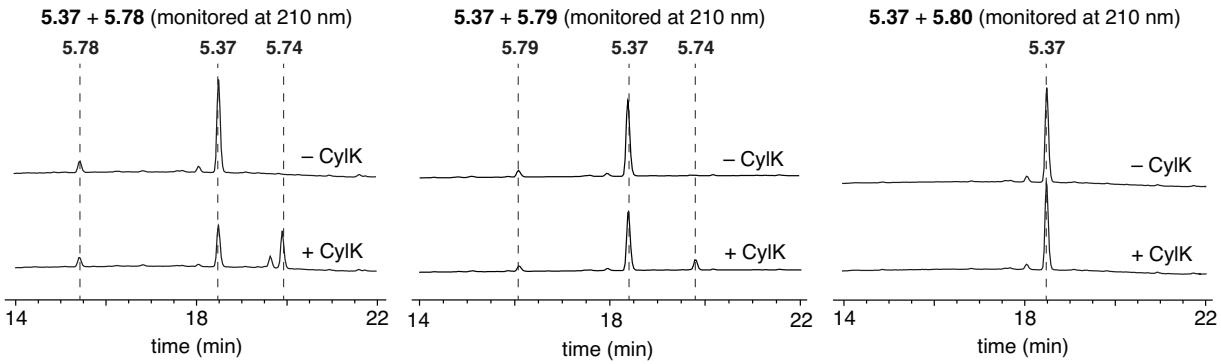
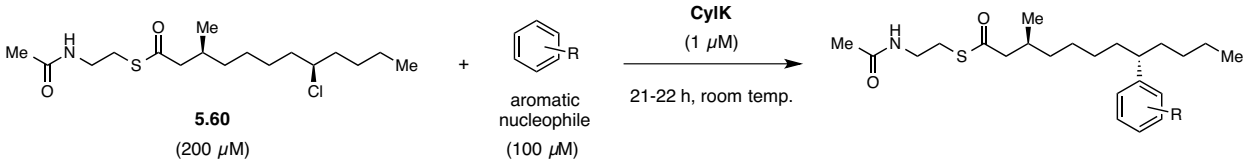


Figure 5.29: HPLC analysis of the CylK-catalyzed resorcinol alkylation assays with non-chloride electrophiles.

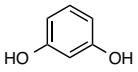
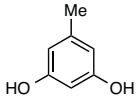
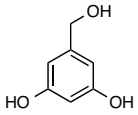
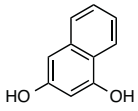
Next, we screened the ability of CylK to use other aromatic nucleophiles. The diastereomerically-enriched (*S,R*)-chlorinated SNAC substrate (**5.60**), which resulted in the best conversion in the alkyl chloride substrate screen, was chosen as the alkyl chloride partner (**Scheme 5.19**). The aromatic nucleophile was the limiting substrate in each of the assays, and the conversion was analyzed using HPLC (**Figures 5.30** and **5.31**). The assay result showed that CylK could use all the nucleophiles tested that contained the resorcinol moiety (**5.75-5.78**, see **Table 5.6**). Even resorcinol (**5.81**), which lacks an alkyl substituent, and 1,3-naphthalenediol (**5.84**), which contains a bulky fused ring system, showed 9% and

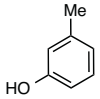
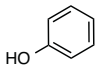
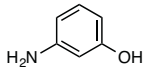
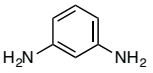
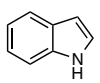
26% conversion, respectively. Aromatic rings containing only one hydroxyl group, including cresol (**5.85**) and phenol (**5.86**), did not react, which indicates that the second hydroxyl group of the resorcinol moiety is important for either binding to the active site or for enhancing nucleophilicity by increasing the electron density of the aromatic ring. No conversion was observed for aromatic nucleophiles in which amine groups replaced one or more of the hydroxyl groups (**5.87** and **5.88**), which might be due to protonation of the basic amino group under the assay condition (pH 8.0). The last nucleophilic heterocycle tested, indole (**5.89**), also showed no conversion. The assay products were also detected by LC-MS, and in all cases, except for the assay with 1,3-napthalene diol (**5.84**), just a single product was detected.



Scheme 5.19: CyK-catalyzed coupling of the (*S,R*)-chlorinated SNAC and various aromatic nucleophiles.

Table 5.6: HPLC conversion (%) of the CylK-catalyzed alkylation with various aromatic nucleophiles.

aromatic nucleophile		conversion
	**5.81**	9 %
	**5.82**	30 %
	**5.83**	7 %
	**5.84**	26 %

aromatic nucleophile		conversion
	**5.85**	0 %
	**5.86**	0 %
	**5.87**	0 %
	**5.88**	0 %
	**5.89**	0 %

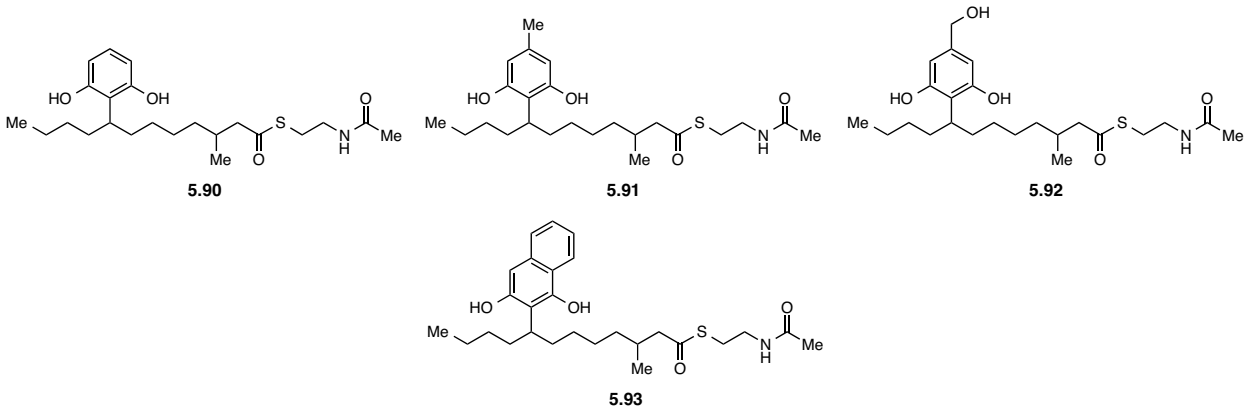


Figure 5.30: Predicted structures of the coupled products of the CylK-catalyzed alkylation of the various aromatic nucleophiles with the product formation observed.

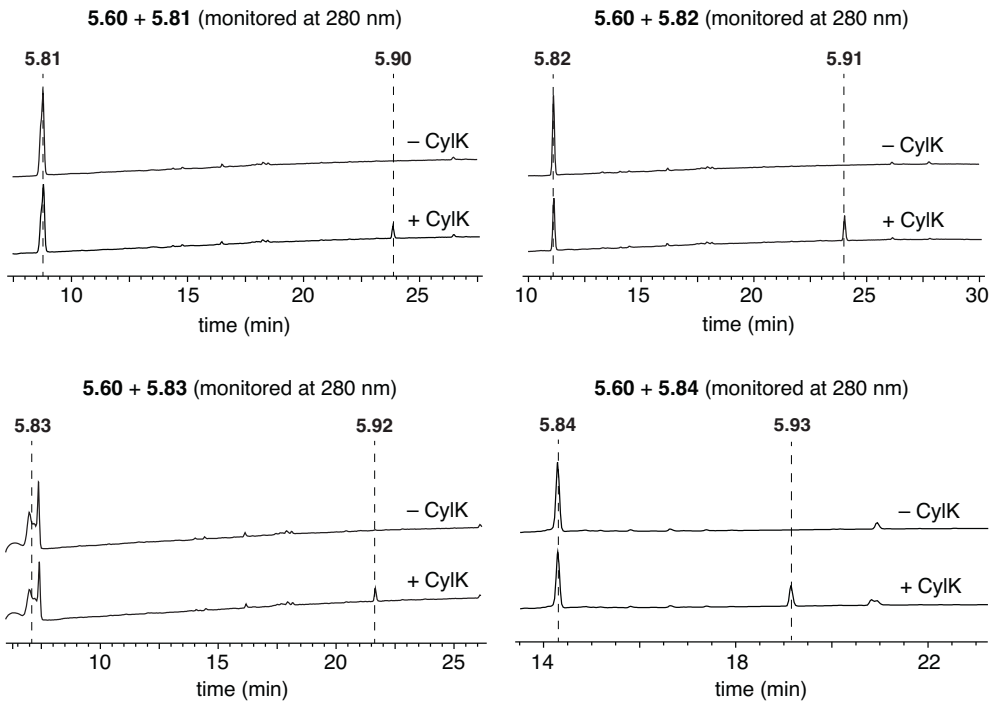
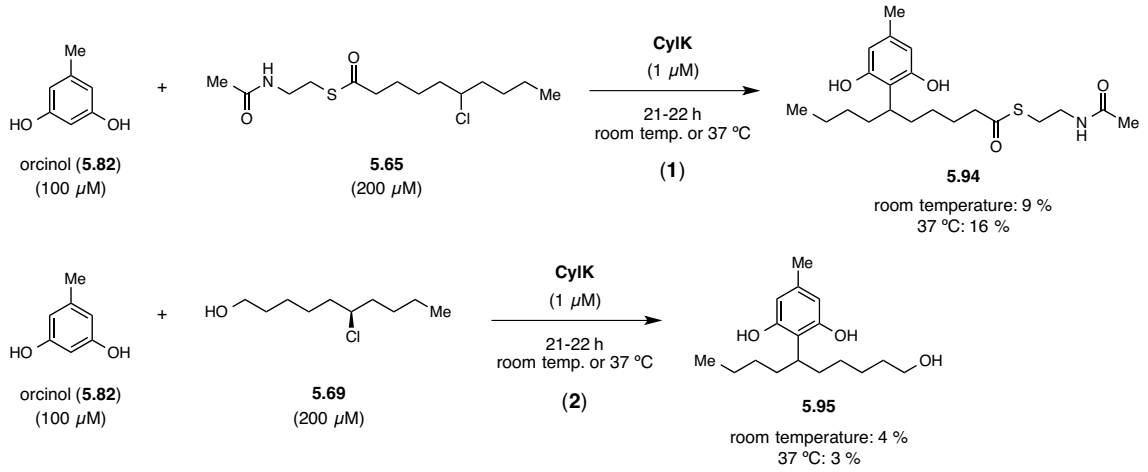


Figure 5.31: HPLC analysis of the CylK-catalyzed alkylation of different aromatic nucleophiles with the product formation observed.

Finally, we tested whether CylK is capable of forming C–C bonds between two simpler resorcinol and alkyl chloride substrates. Based on the previous substrate scope assay results, we decided to test the combinations of orcinol (**5.82**) with the chlorinated decanoyl-SNAC (**5.65**) and the chlorinated decanol (**5.69**, see **Scheme 5.20**). The assays were performed under the same conditions as the previous assays

either at room temperature or 37 ºC. Some conversion was observed for both combinations of substrates by HPLC analysis, demonstrating that CylK is capable of coupling two abbreviated substrates together (**Figure 5.32**). The conversion was slightly better at 37 ºC for the chlorinated decanoyl-SNAC substrate (**5.65**), but equally poor at both temperatures tested for the chlorinated decanol (**5.69**). The result of these assays indicates that CylK is indeed a promiscuous enzyme with a broad substrate scope.



Scheme 5.20: CylK substrate scope assay with orcinol and chloro-decanoyl-SNAC or chloro-decanol.

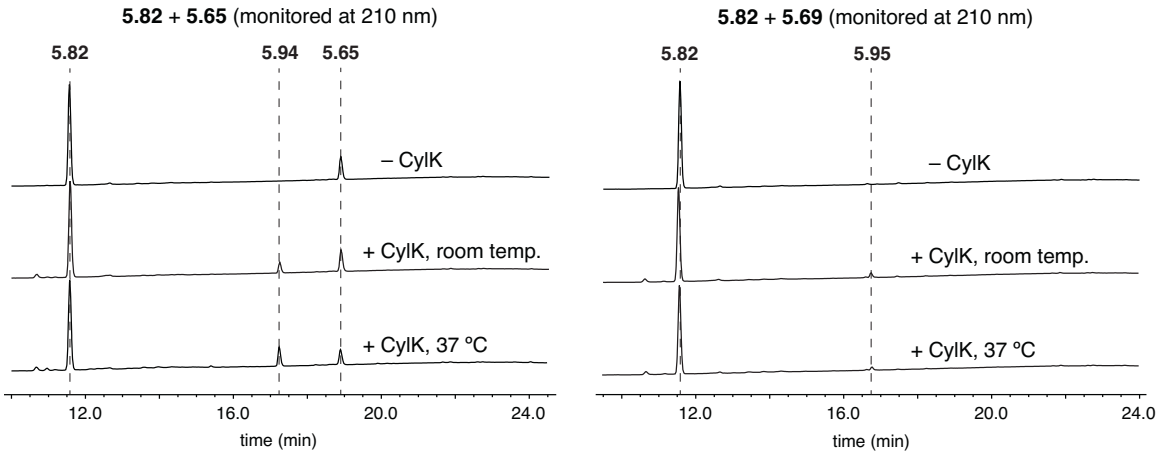


Figure 5.32: HPLC analysis of the CylK assay with orcinol and chloro-decanoyl-SNAC or chloro-decanol.

5.2.8. Bioinformatic search for other biosynthetic gene clusters with CylK homologs

Our investigation of the cylindrocyclophane biosynthesis led to the discovery of the novel C–C bond-forming enzyme, CylK. To determine the distribution of CylK homologs in other bacterial biosynthetic gene clusters, we performed a BLAST search and examined the genomic context of the hits. Interestingly, all of the CylK homologs that appeared to be a part of natural product biosynthetic gene clusters (18 out of 27 hits) are found together with CylC homologs (**Table 5.7**). The hit list includes CylK and CabK from the cylindrocyclophane and the carbamidocyclophane biosynthetic gene clusters from *Cylindrospermum stagnale* PCC 7417[15] and *Nostoc* sp. CAVN2,[17] respectively. In addition, the bartoloside biosynthetic gene cluster in *Synechocystis salina* LEGE 06155[46] also has a CylK homolog (BrtB), but the function of this enzyme in the biosynthesis of the bartolosides is unknown.

Table 5.7: CylK homologs clustered with CylC homologs

CylK homolog accession	Length (%ID, %similarity)	Strain	CylC homolog length (%ID, %similarity)
WP_015207394.1	681 (92/95)	*Cylindrospermum stagnale* PCC 7417	471 (98/99)
CabK (AMB48443.1)	687 (79/89)	*Nostoc* sp. CAVN2	471 (96/97)
WP_058184687.1	768 (43/60)	*Mastigocoleus testarum* BC008	467 (50/69)
WP_012409773.1	861 (40/56)	*Nostoc punctiforme* PCC 73102	456 (49/64)
WP_019489587.1	1740; 46-624 (42/60) 708-1312 (41/56)	*Calothrix* sp. PCC 7103	456 (49/64)
WP_019504293.1	648 (38/53)	*Pleurocapsa* sp. PCC 7319	446 (42/59)
WP_052672355.1	729 (39/57)	Chroococcales cyanobacterium CENA 595	470 (45/63)
WP_051470317.1	735 (37/52)	*Fischerella* sp. PCC 9605	453 (50/66)
ADN17704.1	834 (37/53)	*Cyanothece* sp. PCC 7822	454 (48/66)
WP_015142025.1	749 (36/52)	*Pleurocapsa* sp. PCC 7327	473 (48/62)
WP_016870032.1	604 (48/65)	*Fischerella muscicola* PCC 7417	410 (46/61)
WP_053455401.1	590 (46/63)	*Hapalosiphon* sp. MRB220	410 (46/60)

(Continued)

WP_016864402.1	605 (46/63)	*Fischerella muscicola* SAG 1427-1	405 (45/58)
WP_026719276.1	590 (45/63)	*Fischerella* sp. PCC 9431	410 (46/59)
WP_009459339.1	590 (45/62)	*Fischerella* sp. JSC-11, *Ficherella thermalis* PCC 7521	408 (47/60)
AFZ58481.1	765 (45/64)	*Anabaena cylindrica* PCC 7122	403 (47/61)
BrtB (AKV71847.1)	834 (30/46)	*Synechocystis salina* LEGE 06155	401 (40/55)

All of the biosynthetic gene clusters with CylC and CylK homologs also contain at least one polyketide synthase and either DarAB homologs (**Figure 5.33**) or fatty acid activating enzymes (**Figure 5.34**). DarA and DarB are enzymes that generate dialkylresorcinols through condensation of two acyl-CoA or acyl-ACP substrates (**Scheme 5.21**). DarAB homologs are also found in the bartoloside biosynthetic gene cluster (BrtC and BrtD).[46] DarAB-encoding gene clusters appear to be very conserved across multiple organisms and strains, which may have resulted from horizontal gene transfers. On the other hand, the gene clusters containing the fatty acid activating enzymes (CylA and CylB homologs) are more varied. For instance, the gene cluster in *Pleurocapsa* sp. PCC 7319 contains a type III PKS as well as an □-ketoglutarate-dependent non-heme iron halogenase in addition to CylC and CylK homologs. Many of these gene clusters also encode glycosyltransferases, which indicate that these clusters might be responsible for producing glycolipids.

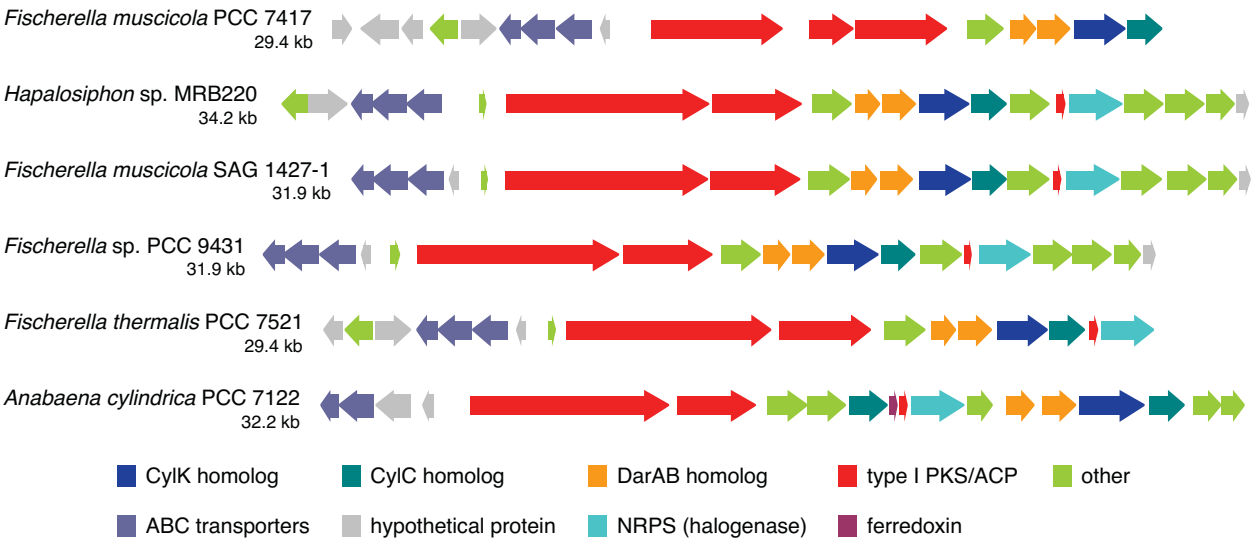
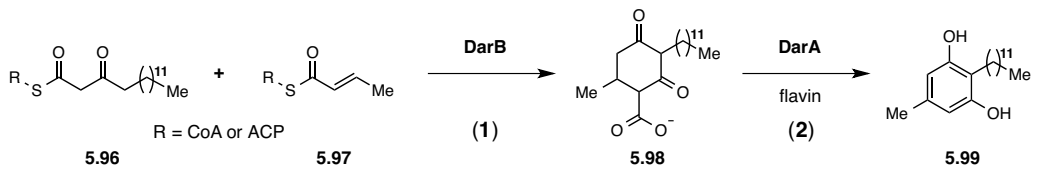


Figure 5.33: Biosynthetic gene clusters with CylC and CylK homologs as well as DarAB homologs.

Figure 5.34: Biosynthetic gene clusters with CylC and CylK homologs as well as fatty acid activating enzymes.



Scheme 5.21: Example of the dialkylresorcinol formation catalyzed by DarA and DarB.

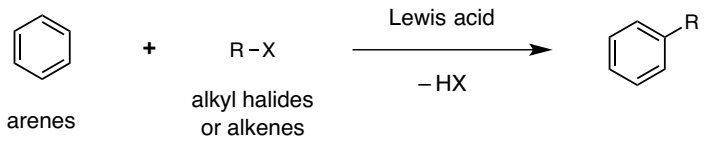
The natural product biosynthetic gene clusters containing CylK and CylC homologs appear to be present only in cyanobacteria. The co-occurrence of these genes in a diverse set of gene clusters suggests that the generation of an alkyl chloride by a CylC homolog followed by a C–C bond formation by a CylK homolog might be a common strategy used in the biosynthetic pathways of cyanobacteria. Characterization of the CylC and CylK homologs in other biosynthetic gene clusters is necessary to determine the distribution of this C–C bond formation logic.

5.3. Conclusions

Investigation of the cylindrocyclophane biosynthesis using a chemically guided approach has led to the discovery of the C–C bond-forming enzyme CylK. By comparing the gene organizations of the four cylindrocyclophane[14-16] and the carbamidocyclophane biosynthetic gene clusters,[17] we identified CylJ and CylK as two candidate enzymes that catalyzes the dimerization step. Previous work on CylJ homologs[24-26] and the observations from feeding studies suggest that CylJ catalyzes *O*-methylation of the resorcinol moiety as a detoxification strategy. CylK is predicted to be a di-domain protein with an N-terminal calcium-binding domain and a C-terminal β-propeller repeat domain. Secondary structural homologs of CylK catalyze enolate chemistry, which suggested that CylK might catalyze the C–C bond formation in the dimerization step using the phenolate of the resorcinol moiety as the nucleophile.

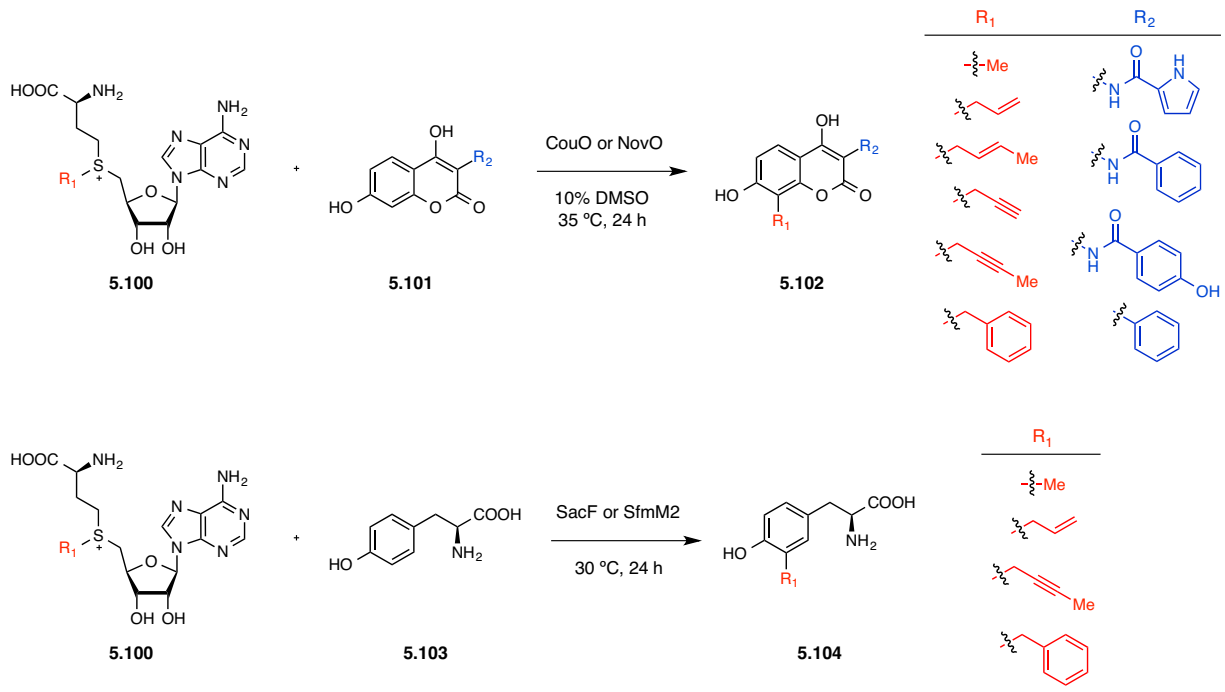
The CylK activity was investigated in the coupled *in vitro* assay with the type III PKS CylI, and both the production of cylindrocyclophane F (**5.51**) as well as an intermediate with one newly-formed C–C bond (**5.52**) were observed. This result is the first experimental evidence supporting the function of CylK as a novel C–C bond-forming enzyme. The production of cylindrocyclophane F from chlorinated substrates

(**5.50** and **5.53**) in both the coupled and the non-coupled assays validated our hypothesis that cylindrocyclophane biosynthesis proceeds through cryptic chlorination to construct the paracyclophane scaffold. Further biochemical assays show that CylK is a calcium-dependent enzyme that catalyzes a stereospecific S_N2-type C–C bond formation. The reaction catalyzed by CylK is analogous to the Friedel–Crafts alkylation used in organic synthesis, which is a Lewis acid-catalyzed electrophilic aromatic substitution of arenes with alkyl halides or alkenes (**Scheme 5.22**).[47] The Friedel–Crafts alkylation has been known since the late 1800s and is very important organic reaction for C–C bond formation.[47]



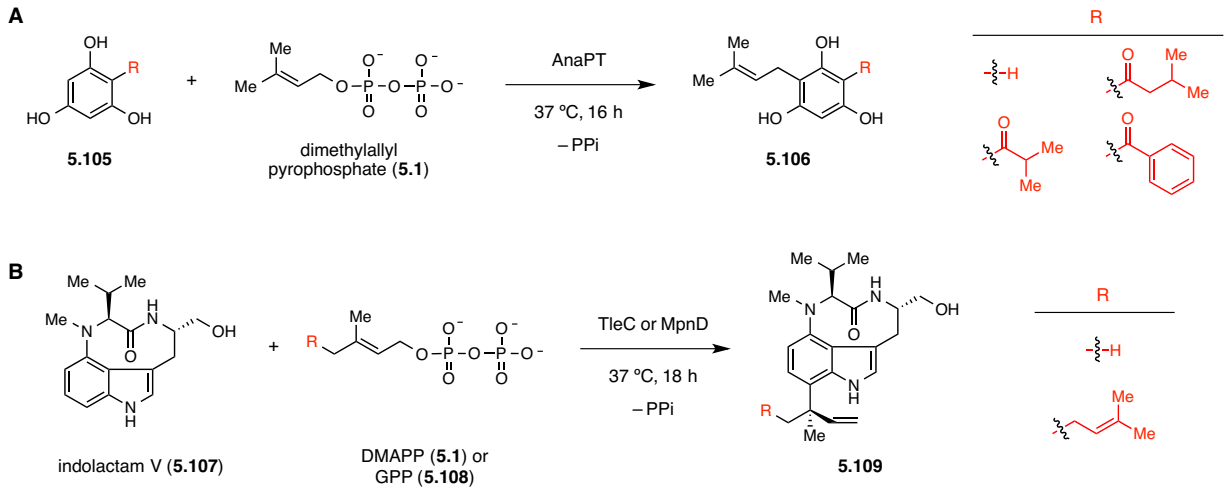
Scheme 5.22: Friedel–Crafts alkylation.

Other enzymes that have been reported to catalyze Friedel–Crafts-type alkylations are SAM-dependent methyltransferases,[9] prenyltransferases[48] and the squalene hopene cyclases (SHCs).[49,50] The potential use of these enzymes as biocatalysts for Fridel–Crafts-type alkylation has been previously investigated. Several methyltransferases have been demonstrated to transfer various alkyl groups to aromatic substrates using unnatural SAM-analogs (**Scheme 5.23**).[51,52] While methyltransferases may be useful for catalyzing alkylations of the aromatic compounds under mild conditions, the practicality of these reactions are limited by the need to synthesize the unnatural SAM-analogs that serve as alkyl donors. Moreover, methyltransferases generally have high specificity for their aromatic substrate, which indicates that a different methyltransferase is required for each aromatic substrate.



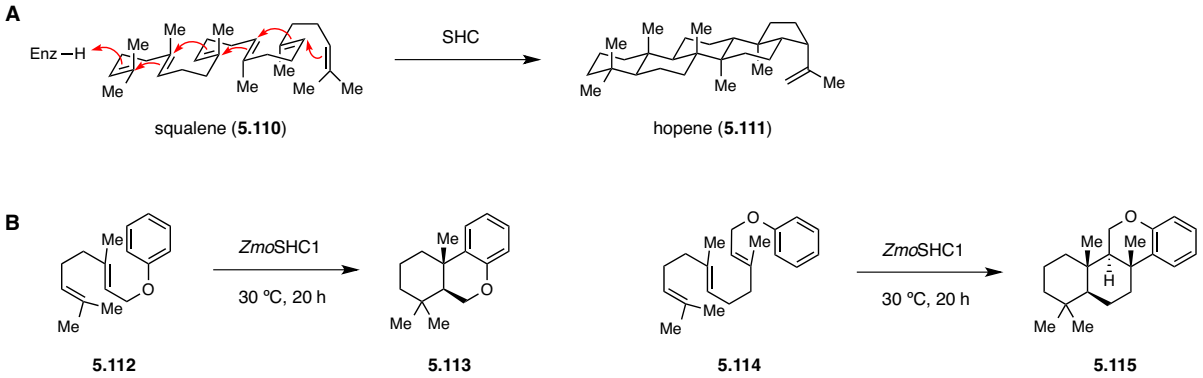
Scheme 5.23: Friedel–Crafts alkylations catalyzed by methyltransferases using unnatural SAM-analogs.

Prenyltransferases are also useful enzymes for installing alkyl substituents on nucleophilic aromatic substrates, and substrate scopes of several prenyltransferases have been reported.[48,53-55] Prenyltransferases can either catalyze 'normal' (**Scheme 5.24A**)[53] or 'reverse' prenylation (**Scheme 5.24B**)[55] to form C–C bond at the C-1 primary carbon or the C-3 tertiary carbon center, respectively. While a diverse set of prenyltransferases that alkylates specific positions on varieties of aromatic substrates is known, the alkyl substituents that can be introduced by these enzymes are limited to isoprene units with diphosphate as the leaving group.[48,53-55] This restraint on alkyl donors limits the use of prenyltransferases as biocatalyts.



Scheme 5.24: Reported substrate scope screens for prenyltransferases. **A**) Examples of 'normal' prenylation. **B**) Examples of 'reverse' prenylation.

Another enzyme that has been reported to catalyze Friedel–Crafts-type alkylation is the squalene hopene cyclase (SHC).[49] SHCs belong to the class II terpene synthases and unlike the class I terpene synthases, these enzymes do not require their substrates to have a pyrophosphate leaving group.[50] Instead, SHCs catalyze C–C bond formation by a cationic polycyclization cascade of the squalene substrate through initial protonation of the terminal isoprene C=C bond (**Scheme 5.25A**).[50] Because SHCs do not require a pyrophosphate leaving group, they have been considered as potentially useful biocatalysts for C–C as well as C–O and C–N bond formation.[50] SHCs tend to be relatively promiscuous, and the SHC from *Zymomonas mobilis* has been demonstrated to catalyze Friedel–Crafts-type alkylation using aromatic substrate (**Scheme 5.25B**).[49] The current use of SHCs, however, is limited to the construction of ring systems through intramolecular C–C bond formation.



Scheme 5.25: Reactions catalyzed by squalene hopene cyclases. **A**) Native reaction catalyzed by SHCs. **B**) Examples of Friedel–Crafts alkylation catalyzed by SHC from *Zymomonas mobilis*.

While various enzymes that catalyze Friedel–Crafts-type alkylations are known, the CylK-catalyzed Friedel–Crafts alkylation is distinct from those of the previously reported enzymes. To the best of our knowledge, CylK is the first enzyme to use alkyl halides as electrophilic partners for the Friedel–Crafts alkylation. The enzyme that uses alkyl halide is beneficial for biocatalysis, since alkyl halides are readily available in organic synthesis. In addition, CylK naturally has a broad substrate scope and such promiscuity is helpful in bioengineering to develop useful biocatalysts through methods such as directed evolution.[56] Although the substrate scope of the nucleophilic partner is currently limited to compounds with the resorcinol moiety, we will attempt to expand the substrate scope and improve reactivity with non-natural substrate pairs through future bioengineering efforts, which will involve structure elucidation and detailed mechanistic studies. This work now sets the stage for engineering biocatalysts for aromatic ring alkylation.

Through combinations of bioinformatics, feeding experiments and biochemical assays in a chemically guided approach, we succeeded in identifying all the enzymes that are required for the construction of the core [7.7]paracyclophane scaffold in the cylindrocyclophanes. Our elucidation of the cylindrocyclophane biosynthesis led to the discovery of the novel halogenase CylC and the C–C bond-forming enzyme CylK that catalyzes a Fridel–Crafts alkylation. The co-distribution of CylC and CylK homologs in many strains

of cyanobacteria indicate that the C–C bond formation through cryptic halogenation might be a common biosynthetic strategy used by cyanobacteria.

5.4. Materials and Methods

Syntheses of the diastereomeric mixtures of resorcinol (**5.37**), the β-ketoacyl-SNAC substrates (**5.49**), and the decanoyl-SNAC substrates (**5.65**, **5.78** and **5.80**) are described in **Chapter 2.4.14–16**. The synthesis of the chlorinated β-ketoacyl-SNAC and the acyl-SNAC substrates (**5.50** and **5.61**) are described in **Chapter 4.4.17**.

The expression and purification of N-His_6-CylI and C-His_6-$CylH_{PKS}$ are described in **Chapter 2.4.6** and **Chapter 3.4.3**, respectively.

5.4.1. Materials and General Methods

All chemical reagents were purchased from Sigma–Aldrich, unless otherwise indicated. Luria-Bertani Lennox (LB) medium was purchased from EMD Millipore or Alfa Aesar and IPTG was purchased from Teknova. All NMR solvents were purchased from Cambridge Isotope Laboratories. Primers were purchased from Integrated DNA Technologies. Thermocycling was carried out in a C1000 Gradient Cycler (Bio-Rad). DNA polymerases, restriction enzymes and T4 ligase for cloning were purchased from New England BioLabs. Recombinant plasmid DNA was purified with a QIAprep Kit from Qiagen. Gel extraction of DNA fragments and restriction endonuclease clean up were performed using an Illustra GFX PCR DNA and Gel Band Purification kit from GE Healthcare. DNA sequencing was performed by Beckman Coulter Genomics (Danvers, MA). Nickel-nitrotriacetic acid-agarose (Ni-NTA) resin was purchased from Qiagen and Thermo Scientific, and *Strep*-Tactin resin was purchased from IBA. SDS-PAGE gels were purchased from Bio-Rad.

DNA concentrations were determined using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific). Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm. Gel filtration FPLC was

performed on a BioLogic DuoFlow Chromatography System (Bio-Rad) equipped with a Superdex column, 200 10/300 GL (GE Healthcare). Bovine thyroglobulin (670 kDa), bovine □-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa), and vitamin B_{12} (1350 Da) were used as molecular weight markers (Bio-Rad #151-190). The molecular weights of the proteins analyzed by gel filtration were calculated from their elution volume, using a linear relationship between log (molecular weight) and ve/vo (elution volume/void volume). Protein concentrations were determined using a NanoDrop 2000 UV-Vis Spectrophotometer for CStrep-CylK (□ = 148,740 $M^{-1}cm^{-1}$), and the extinction coefficients were calculated using ExPASy ProtParam tool.

High-resolution mass spectral data and the spectrophotometric assay data were obtained in the Small Molecule Mass Spectrometry Facility, FAS Division of Science. Mass spectral data were obtained on an Agilent 1290 Infinity UHPLC system (Agilent Technologies) and a maXis impact UHR time-of-flight mass spectrometer system (Bruker Daltonics Inc) equipped with an electrospray ionization (ESI) source. High-resolution mass spectral (HRMS) data for the synthetic compounds were obtained on an Agilent 6210 TOF. The capillary voltage was set to 4.5 kV and the end plate offset to –500 V, the drying gas temperature was maintained at 190 ^{o}C, with a flow rate of 8 L/min and a nebulizer pressure of 21.8 psi. The liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with 50% H_2O and 50% acetonitrile as solvent.

Analytical HPLC was performed on a Dionex Ultimate 3000 instrument (Thermo Scientific). Proton nuclear magnetic resonance (^{1}H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were obtained on a Varian Inova-500 (500 MHz, 125 MHz) NMR spectrometer in the Magnetic Resonance Laboratory in the Harvard University Department of Chemistry and Chemical Biology. Chemical shifts are reported in parts per million downfield from tetramethylsilane using the solvent resonance(s) as internal standard for ^{1}H ($CDCl_3$, δ_H = 7.26 ppm or CD_3OD, δ_H = 3.35, 4.90 ppm) and ^{13}C ($CDCl_3$, δ_C = 77.2 ppm). Data are reported as follows: chemical shift, multiplicity (s = singlet, d =

doublet, t = triplet, m = multiplet), integration, and coupling constant. NMR spectra were visualized using iNMR Version 5.3.6 (Mestrelab Research).

5.4.2. HHpred search and generation of homology models for CylJ and CylK

CylJ and CylK sequences were used as queries to search for structural homologs using HHpred.[22] The protein from the PDB database with the best hit for CylJ was isoprenylcysteine carboxyl methyltransferase (ICMT, PDF file 4A2N).[23] CylK appeared to have two domains, and one of the best hits for the N-terminal and C-terminal domains were the R-modules of the C-5 alginate epimerases AlgE4/6 (PDB files 2ML1, 2ML3 and 2AGM)[35] and virginiamycin B lyase (PDB file 2Z2N)[36], respectively. A homology model of CylJ was generated by Modeller (Max-Plank Institute for Developmental Biology) using the structure of ICMT (4A2N) as template. Similarly, a homology model of CylK was generated using both the R-module of AlgE6 (2ML3) and virginiamycin B lyase (2Z2N) as templates. The resulting PDB files were aligned with the template PDB files using MacPyMOL version 1.7 (Schrödinger Inc.).

5.4.3. Cloning of *cylK*

Table 5.8. Oligonucleotides used for cloning *cylK* (restriction sites underlined).

Primer name	Sequence (5' to 3')	Target
CylK-F-BsaI-pPR-IBA1	GTGATCGGTCTCTAATGAAAAAAAACAAAAAAAC	C-Strep-CylK
CylK-R-BsaI-pPR-IBA1	AGTCACGGTCTCAGCGCTCCCTAAATTCAAAAAAC	

CylK was PCR amplified from *Cylindrospermum licheniforme* ATCC 29412 genomic DNA using the primers shown in **Table 5.8**. Both forward and reverse primers have a *Bsa*I restriction site. The PCR mixture contained Phusion High-Fidelity PCR Master Mix, 50 ng DNA template, and 25 pmoles of each primer in a total volume of 50 µL. The thermocycling was carried out using the following parameters: denaturation for 30 sec at 98 ºC, followed by 35 cycles of 10 sec at 98 ºC, 30 sec at 56 ºC, 2 min at 72 ºC,

and a final extension time of 10 min at 72 ºC. PCR products were analyzed by agarose gel electrophoresis with ethidium bromide staining, pooled and purified.

Amplified fragments were digested with *Bsa*I for 2.5 h at 37 ºC. Digests contained 4 µL water, 3 µL Cutsmart buffer (10x), 20 µL of PCR product, 3 µL of *Bsa*I (10 U/µL). Restriction digests were purified directly using agarose gel electrophoresis; gel fragments were further purified using the Illustra GFX kit. The digests were ligated into linearized pPR-IBA1 expression vector using T4 DNA ligase. Ligations were incubated at 16 ºC overnight and contained 3 µL water, 1 µL T4 Ligase Buffer (10x), 1 µL digested vector, 3 µL digested insert DNA, and 2 µL T4 DNA Ligase (400 U/µL). 5 µL of the ligation sample was used to transfer a single tube of chemically competent *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pPR-IBA1-*cylK* construct was confirmed by sequencing of the purified plasmid DNA.

5.4.4. Expression and purification of CylK

The pPR-IBA1-*cylK* vector was transformed into chemically competent *E. coli* Rosetta (DE3) cells (Invitrogen), and the resulting colonies were used directly to prepare a 50 mL starter culture in LB medium supplemented with 50 µg/mL ampicillin and 34 µg/mL chloramphenicol. The starter culture was incubated overnight at 37 ºC. Overnight cultures were diluted 1:100 into 2 L LB medium containing ampicillin and chloramphenicol. The cultures were incubated at 37 ºC with 190 rpm shaking. At OD_{600} = 0.25, the cultures were cooled to 15 ºC and further incubated with 175 rpm shaking. When the cultures reached OD_{600} = 0.6–0.7, 5 mM $CaCl_2$ was added to each culture, and the protein expression was induced with 100 µM IPTG. The cultures were incubated for 16 h at 15 ºC with 175 rpm shaking.

Cells from 2 L of culture were pelleted by centrifugation (6,720 □*g* for 15 min) and resuspended in 50 mL lysis buffer (20 mM HEPES, 500 mM NaCl, 20 mM $CaCl_2$, 10 mM $MgCl_2$, pH 8) supplemented with EDTA-free Pierce Protease Inhibitor Tablets (Thermo Fisher). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 8,000–10,000 psi, and the lysate was clarified by centrifugation (28,900 □*g* for 30 min). The supernatant was incubated with 2 mL of *Strep*-Tactin resin for

2 h at 4 ºC. The mixture was centrifuged (3,200 ×*g* for 5 min) and the unbound fraction was discarded. The *Strep*-Tactin resin was resuspended in 5 mL of wash buffer (lysis buffer lacking detergent), loaded into a glass column, and washed with 8 mL of wash buffer. Protein was eluted from the column in elution buffer (lysis buffer containing 2.5 mM desthiobiotin) in multiple 1 mL fractions. SDS-PAGE analysis (4–15% Mini-PROTEAN TGX precast gel) was employed to determine the presence and purity of protein in each fraction. Fractions containing the desired protein were combined and dialyzed twice against 2 L of dialysis buffer (20 mM EPPS, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, pH 8.0). The dialyzed protein solution was concentrated using a Spin-X UF 20 mL centrifugal concentrator with a 30,000 MWCO membrane (Corning), and the concentrated protein solutions were frozen with liquid nitrogen for storage at –80 ºC. This procedure afforded 0.1 mg/L of crude C-Strep-CylK.

C-Strep-CylK was further purified by gel filtration FPLC. A ~20 μM solution of C-Strep-CylK was purified by gel filtration. The protein was eluted in buffer containing divalent metals (20 mM HEPES, pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10% glycerol) or buffer lacking divalent metals (20 mM HEPES, pH 8.0, 50 mM NaCl, 10% glycerol) at 0.25 mL/min using Superdex 200 30/100 GL column. A solution of molecular weight markers was analyzed under the same conditions. Fractions containing C-Strep-CylK was combined, concentrated using a Spin-X UF 20 mL centrifugal concentrator with a 30,000 MWCO membrane (Corning), and frozen with liquid nitrogen for storage at –80 ºC.

5.4.5. Coupled CylI and CylK assay using β-ketoacyl-SNAC substrates

An assay mixture contained 100 μM substrate (**5.49** or **5.50**), 5 μM N-His_6-CylI, 1 μM C-Strep-CylK, 400 μM malonyl-CoA, 10 mM $CaCl_2$, 10 mM $MgCl_2$ and 100 mM NaCl in 20 mM HEPES buffer at pH 8 in a final assay volume of 100 μL. An aqueous solution containing HEPES buffer, $CaCl_2$, and malonyl-CoA was prepared first. The mixture was vortexed, the enzymes were added, and the reaction was initiated by the addition of the substrate. The reaction was mixed gently and incubated at room temperature. The 20 mM HEPES buffer at pH 8 containing only 100 mM NaCl was used for the metal-dependency assays with or without 1 mM EDTA. After 5 min, 15 min, 30 min, and 1 h, 20 μL aliquots

were removed from the reaction mixture, added to 40 µL of ice-cold 1:1 acetonitrile/methanol solution, vortexed, incubated on ice for 10 min and centrifuged (16,100 □*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC (10 µL injection volume) on a Hypersil Gold C18 column (3.0 x 150 mm, Thermo Scientific) with a guard column at a flow rate of 0.5 mL/min. HPLC assay conditions were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 16 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 10 min (solvent A = water, solvent B = acetonitrile). The LC-MS analysis of the CylI and CylK coupled assays was performed on a Maxis Impact UHR time-of-flight mass spectrometer system using the same LC condition as described above, except on a Gemini C18 column (5 µM, 4.6 x 50 mm) using different solvent system (solvent A = 95:5 water/acetonitrile + 10 mM NH_4OH, solvent B = 2:98 water/acetonitrile + 10 mM NH_4OH). The LC-MS analysis was performed in negative ion mode, and the MS/MS analysis was performed at 35 eV.

5.4.6. Preparative scale CylI assay and isolation of chlorinated resorcinol

To structurally characterize the CylI assay product using the chlorinated □-ketoacyl-SNAC substrate **5.50** and to obtain a pure substrate for CylK assay, 38 □ 1 mL assays were set up. The assay mixtures contained 20 mM HEPES buffer at pH 8.0 containing 100 mM NaCl, 300 µM malonyl-CoA, 100 µM substrate and 5 µM N-His_6-CylI. After 2 h at room temperature, the reactions were combined into a separatory funnel and extracted with ethyl acetate (3 □ 30 mL). The combined organic layer was dried over sodium sulfate, filtered, and concentrated *in vacuo*. The crude organic extract was purified by flash chromatography on silica gel using a gradient of 20-30% ethyl acetate in hexanes to obtain chlorinated resorcinol **5.53** (0.6 mg, 0.002 mmol, 51% yield). Proton, COSY and HSQC NMR experiments were performed in d_4-methanol using a symmetrical NMR tube susceptibility matched to methanol (Shigemi, Inc). TLC: R_f = 0.69 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{18}H_{28}ClO_2^-$ $[M–H]^-$, 311.1783; found 311.1789. ^{1}H-NMR (600 MHz; CD_3OD): δ 6.15 (d, 2H, J = 2.1 Hz, aromatic C**H**), 6.13 (t, 1H, J = 2.1 Hz, aromatic C**H**), 3.94 (m, 1H, C**H**Cl), 2.50 (dd, 1H, J = 13.2, 6.5 Hz, benzylic C$\mathbf{H}_2$),

2.26 (dd, 1H, *J* = 13.0, 8.0 Hz, benzylic C**H**$_2$), 1.81-1.64 (m, 4H, CHClC**H**$_2$ and C**H**), 1.59-1.29 (m, 9H, C**H**$_2$), 1.18 (m, 1H, C**H**$_2$), 0.97 (t, 3H, *J* = 7.2 Hz, C**H**$_3$), 0.90 (d, 3H, *J* = 6.6 Hz, CHC**H**$_3$).

5.4.7. CylK assay with chlorinated resorcinol

An assay mixture contained 200 μM substrate (**5.53**), 0.5 μM C-Strep-CylK, 10 mM $CaCl_2$, 10 mM $MgCl_2$ and 100 mM NaCl in 20 mM HEPES buffer at pH 8 in a final assay volume of 100 μL. An aqueous solution containing HEPES buffer and $CaCl_2$ was prepared first. The mixture was vortexed, the enzyme was added, and the reaction was initiated by the addition of the substrate. The reaction was mixed gently and incubated at room temperature. After 5 min, 10 min, 15 min, 30 min, and 1 h, 20 μL aliquots were removed from the reaction mixture, added to 40 μL of ice-cold 1:1 acetonitrile/methanol solution, vortexed, incubated on ice for 10 min and centrifuged (16,100 □*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC (10 μL injection volume) on a Hypersil Gold C18 column (3.0 x 150 mm, Thermo Scientific) with a guard column at a flow rate of 0.5 mL/min. HPLC assay conditions were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 16 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 10 min (solvent A = water, solvent B = acetonitrile).

5.4.8. EDTA-treatment and purification of CylK by gel filtration FPLC

C-Strep-CylK (5 μM) that has been purified by gel filtration FPLC in buffer lacking divalent metals was incubated with 2 mM EDTA in 20 mM HEPES buffer containing 50 mM NaCl and 10% glycerol at pH 8.0 in a final volume of 200 μL. The sample was left on ice for 10 min, centrifuged at 16,100 □*g* for 10 min, and purified by gel filtration FPLC. The protein was eluted in buffer lacking divalent metals (20 mM HEPES, pH 8.0, 50 mM NaCl, 10% glycerol) at 0.25 mL/min using Superdex 200 30/100 GL column. A solution of molecular weight markers was analyzed under the same conditions. Fractions containing EDTA-treated C-Strep-CylK was combined, concentrated using a Spin-X UF 20 mL centrifugal concentrator with a 30,000 MWCO membrane (Corning), and frozen with liquid nitrogen for storage at –80 ºC.

5.4.9. Metal dependency assay using EDTA-treated CylK

A typical assay mixture contained 200 μM substrate **5.53**, 0.5 μM EDTA-treated C-Strep-CylK, in 100 mM NaCl in 20 mM HEPES buffer at pH 8 in a final assay volume of 50 μL. For screening the dependency of CylK on different divalent metals, $CaCl_2$, $MgCl_2$, $MnCl_2$, $ZnCl_2$ or $(NH_4)_2Fe(SO_4)_2$ was added at 10 mM final concentration. An aqueous solution containing HEPES buffer and $CaCl_2$ was prepared first. The mixture was vortexed, the enzyme was added, and the reaction was initiated by the addition of the substrate. The reaction was mixed gently and incubated at room temperature. After 10 min, 30 min, and 1 h, 15 μL aliquots were removed from the reaction mixture, added to 30 μL of ice-cold 1:1 acetonitrile/methanol solution, vortexed, incubated on ice for 10 min and centrifuged (16,100 □*g* for 10 min at 4 ºC).

To test how the presence of other divalent metals affects the CylK activity in presence of Ca^{2+}, another set of assay was performed. A typical assay contained 100 μM substrate (**5.53**), 0.5 μM EDTA-treated C-Strep-CylK, 100 mM NaCl, 10 mM $CaCl_2$ and 10 mM of another divalent metal solution in 20 mM HEPES buffer at pH 8 in a final volume of 50 μL. After 10 min and 30 min, 20 μL aliquots were removed from the reaction mixture, added to 40 μL of ice-cold 1:1 acetonitrile/methanol solution, vortexed, incubated on ice for 10 min and centrifuged (16,100 □*g* for 10 min at 4 ºC).

The assay samples were analyzed by HPLC (10–20 μL injection volume) on a Hypersil Gold C18 column (3.0 x 150 mm, Thermo Scientific) with a guard column at a flow rate of 0.5 mL/min. HPLC assay conditions were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 16 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 10 min (solvent A = water, solvent B = acetonitrile).

5.4.10. ICP analysis of CylK

The C-Strep-CylK sample for an ICP-MS analysis was expressed and purified following the procedure described in section **5.3.4** with modification in the dialysis step. The CylK sample was dialyzed against 2

 2 L of dialysis buffer lacking divalent metals and glycerol (20 mM HEPES, pH 8, 50 mM NaCl). The samples were then concentrated using a Spin-X UF 20 mL centrifugal concentrator with a 30,000 MWCO membrane (Corning), and the concentrated protein sample was further purified by gel filtration FPLC in the same buffer as the dialysis buffer. The fractions containing C-Strep-CylK were pooled, concentrated (concentration = 0.7 mg/mL, volume = 200 µL), and lyophilized to dryness. The same volume of dialysis buffer was lyophilized as negative control. The sample containing lyophilized C-Strep-CylK and the negative control sample containing just the lyophilized buffer were submitted for ICP-MS analysis to the Microanalysis Laboratory at the University of Illinois Urbana-Champaign. The analysis was performed for Mg, Ca, Mn, Fe and Zn, and the quantity bound to C-Strep-CylK was calculated by subtracting the concentration of each metal found in the buffer negative control from that of the CylK sample.

5.4.11. Chiral HPLC analysis of chlorinated resorcinol

The diastereomers of the chlorinated resorcinol **5.53** in CylK assay samples (injection volume = 30 µL) were analyzed by HPLC using chiral Amylose-1 column (4.6 x 100 mm, Phenomenex) at a flow rate of 1 mL/min. HPLC assay conditions were: 50% B in solvent A for 2 min, a gradient increasing to 95% B in solvent A over 15 min, a gradient increasing to 100% B in solvent A over 1 min, 100% B for 10 min, a gradient decreasing to 50% B in solvent A over 0.5 min, 50% B for 5.5 min (solvent A = water with 10 mM ammonium acetate, solvent B = acetonitrile).

5.4.12. Chiral HPLC analysis of acyl-SNAC substrate for CylH

The diastereomers of the chlorinated acyl-SNAC substrate **5.61** in CylK assay samples (injection volume = 30 µL) were analyzed by HPLC using chiral Amylose-2 column (4.6 x 100 mm, Phenomenex) at a flow rate of 1 mL/min. HPLC assay conditions were: 40% B in solvent A for 2 min, a gradient increasing to 50% B in solvent A over 15 min, 50% B for 13 min, a gradient decreasing to 40% B in solvent A over 1 min, 40% B for 9 min (solvent A = water with 10 mM ammonium acetate, solvent B = acetonitrile). The diastereomeric ratios of the synthesized substrates were determined from the peak areas of the HPLC spectra (monitored at 230 nm).

Diastereomeric ratio of **5.59**: (3*R*,8*S*):(3*S*,8*S*) = 2:98; peaks corresponding to (3*R*,8*R*) and (3*S*,8*R*) were not observed.

Diastereomeric ratio of **5.60**: (3*R*,8*S*):(3*R*,8*R*):(3*S*,8*R*) =3:3:94; peak corresponding to (3*S*,8*S*) and (3*S*,8*R*) was not observed.

5.4.13. Coupled CylH, CylI and CylK assays

A solution of *holo* C-His$_6$-CylH$_{PKS}$ was prepared prior to the assay by incubating 50 μM *apo* C-His$_6$-CylH$_{PKS}$, 5 μM Sfp, 250 μM CoA and 1 mM $MgCl_2$ in 50 mM potassium phosphate buffer at pH 8.0 in 80 μL volume for 1 h at room temperature. The chlorinated resorcinol was generated *in situ* by incubating 100 μM of the chlorinated acyl-SNAC substrate (**5.61**, **5.59** or **5.60**), 20 μM *holo* C-His$_6$-CylH$_{PKS}$, 5 μM N-His$_6$-CylI and 500 μM malonyl-CoA in 100 mM potassium phosphate buffer at pH 8.0 in 50 μL volume for 1 h at room temperature. Then, 0.5 μM C-Strep-CylK and 180 μM malonyl-CoA were added and the reaction was incubated further at room temperature for 30 min or 2 h. The assays were quenched with 2 □ volume of ice-cold 1:1 acetonitrile/methanol solution and incubated on ice for 10 min. The samples were centrifuged (16,100 □*g* for 10 min at 4 ºC) and the supernatant was analyzed by HPLC (10–20 μL injection volume) on a Hypersil Gold C18 column (3.0 x 150 mm, Thermo Scientific) with a guard column at a flow rate of 0.5 mL/min. HPLC assay conditions were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 16 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 10 min (solvent A = water, solvent B = acetonitrile).

5.4.14. CylK substrate scope screen

A typical assay mixture contained 100 μM resorcinol (**5.37**) or other aromatic nucleophile (**5.81**-**5.89**), 200 μM SNAC substrate (**5.60**), 1 μM C-Strep-CylK, 10 mM $CaCl_2$, 10 mM $MgCl_2$ and 100 mM NaCl in 20 mM HEPES buffer at pH 8 in a final assay volume of 30 μL. An aqueous solution containing HEPES buffer and $CaCl_2$ was prepared first. The mixture was vortexed, the enzyme was added, and the reaction was initiated by the addition of the substrate. The reaction was mixed gently and incubated at room

temperature or 37 ºC for 21–22 h. Each assay was quenched with 60 µL of ice-cold 1:1 acetonitrile/methanol solution, vortexed, incubated on ice for 10 min and centrifuged (16,100 □*g* for 10 min at 4 ºC). The supernatant was analyzed by HPLC (10 µL injection volume) on a Hypersil Gold C18 column (3.0 x 150 mm, Thermo Scientific) with a guard column at a flow rate of 0.5 mL/min. The conversion of each reaction was calculated from the peak areas of the resorcinol substrate and the coupled product (monitored at 210 or 280 nm).

HPLC assay conditions for all assays performed with **5.37** were: 20% B in solvent A for 2 min, a gradient increasing to 100% B in solvent A over 16 min, 100% B for 5 min, a gradient decreasing to 20% B in solvent A over 1 min, 20% B for 10 min (solvent A = water, solvent B = acetonitrile).

HPLC assay conditions for all assays performed with other aromatic nucleophiles (**5.81**–**5.89**) were: 0% B in solvent A for 4 min, a gradient increasing to 100% B in solvent A over 16 min, 100% B for 5 min, a gradient decreasing to 0% B in solvent A over 1 min, 0% B for 8 min (solvent A = water, solvent B = acetonitrile).

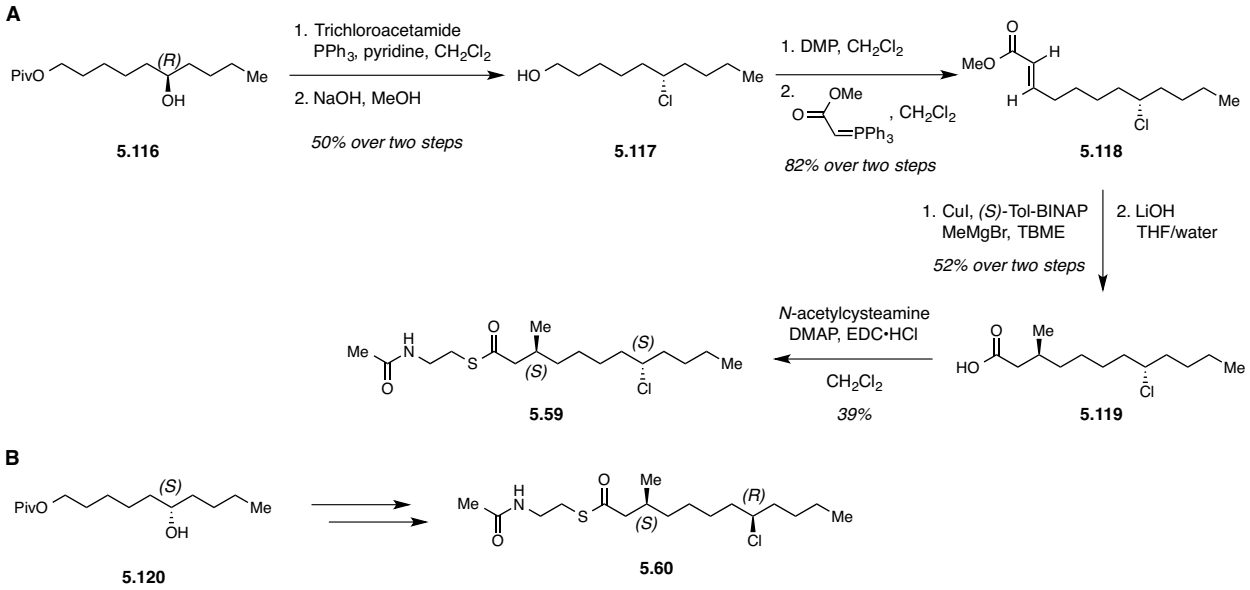
The LC-MS analysis of the substrate scope assays was performed on a Maxis Impact UHR time-of-flight mass spectrometer system using the same LC condition as the HPLC assay described above, except at a flow rate of 0.2 ml/min on a Kinetex C18 column (2.6 µm, 100Å, 2.1 x 150 mm) using different solvent system (solvent A = water, solvent B = acetonitrile with 10 mM NH_4OH). The analysis was performed in negative ion mode, with MS/MS analysis performed at 35, 55 and 75 eV.

5.4.15. Bioinformatic analysis of the genomic context of the CylK homologs

CylK homologs were identified using BLASTp search, and the corresponding nucleotide sequences of the hits were examined to determine the genomic context of the top 100 hits. The % identity and % similarity values of CylK homologs were obtained from BLASTp search results on NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For the homologs that appeared to be part of natural product biosynthetic gene clusters, the nucleotide sequences of the gene clusters were downloaded. The open

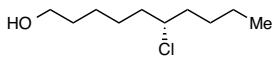
reading frames (ORFs) in each gene cluster were identified using SnapGene Viewer (GSL Biotech LLC). Each ORF was annotated by performing BLASTp search using the translated amino acid sequence as a query. The % identity and % similarity values of CylC homologs were obtained by performing ClustalW multiple sequence alignment of the homologs with CylC using Geneious 7.1.6 (Biomatters).

5.4.16. Synthesis of diastereomerically enriched chlorinated acyl-SNAC CylH substrates



Scheme 5.26: Synthesis of the chlorinated acyl-SNAC substrates for CylH.

5.116 and **5.120** were synthesized following the procedure described in **Chapter 2.4.18**.

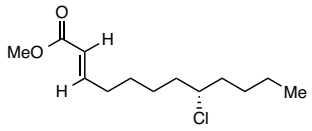


5.117

(*S*)-6-Chlorodecanol (5.117): (*R*)-6-hydroxydecyl pivalate **5.116** (353 mg, 1.37 mmol) was dissolved in anhydrous dichloromethane (7 mL). To the reaction mixture, pyridine (330 µL, 4.10 mmol), triphenylphosphine (537 mg, 2.05 mmol), and 2,2,2-trichloroacetamide (333 mg, 2.05 mmol) were added, and the reaction mixture was stirred at room temperature overnight. The reaction mixture was quenched with water (5 mL) and washed with brine (3 □ 10-mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash

chromatography, eluting with 10-15% ethyl acetate in hexanes to afford (*S*)-6-chlorodecyl pivalate. The crude product co-eluted with excess triphenylphosphine reagent and was carried onto the next reaction without further purification. TLC: R_f = 0.77 (silica gel, 4:1 hexanes/ethyl acetate). ^{1}H-NMR (500 MHz; $CDCl_3$): δ 4.04 (t, 2H, *J* = 6.6 Hz, OC**H**$_2$), 3.86 (m, 1H, C**H**Cl), 1.76-1.21 (m, 14H, C**H**$_2$), 1.18 (s, 9 H, C(C**H**$_3$)$_3$), 0.90 (t, *J* = 7.2 Hz, 3H, C**H**$_3$).

(*S*)-6-chlorodecyl pivalate (1.37 mmol) was dissolved in methanol (10 mL) and NaOH solution (10 M, 3 mL) was added. The reaction mixture was stirred overnight at room temperature. The reaction mixture was quenched with a saturated aqueous solution of ammonium chloride (10 mL) and extracted with diethyl ether (3 × 10 mL). The combined organic layers were washed with a saturated aqueous solution of sodium bicarbonate (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was further purified by flash chromatography, eluting with 15-20% ethyl acetate in hexanes to afford (*S*)-6-chlorodecanol **5.117** (132 mg, 0.68 mmol, 50%). TLC: R_f = 0.30 (silica gel, 4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{10}H_{22}ClO^+$ $[M+H]^+$, 193.1354; found 193.1341. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.89 (m, 1H, C**H**Cl), 3.64 (t, 2H, *J* = 6.6 Hz, COC**H**$_2$), 1.81-1.42 (m, 14H, C**H**$_2$), 0.90 (t, 3H, *J* = 7.2 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 64.2, 62.8, 38.4, 38.3, 32.6, 28.7, 26.3, 25.3, 22.3, 14.0.

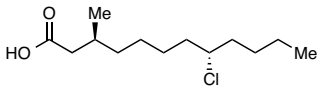


5.118

Methyl (*S*)-8-chlorododec-2-enoate (5.118): (*S*)-6-Chlorodecanol **5.117** (173 mg, 0.90 mmol) was dissolved in anhydrous dichloromethane (9 mL) and cooled to 0 ºC. To the reaction mixture, Dess-Martin periodinane (685 mg, 1.61 mmol) was added, and the reaction mixture was slowly warmed to room temperature. After 30 min, the reaction mixture was cooled to 0 ºC and quenched with a saturated aqueous solution of sodium thiosulfate (10 mL). The reaction mixture was diluted with a saturated aqueous solution of sodium bicarbonate (10 mL) and stirred at room temperature for 30 min. The

quenched reaction mixture was extracted with dichloromethane (3 × 10 mL). The combined organic layers were washed with brine (30 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude (*S*)-6-chlorodecanal was used directly in the next reaction without further purification. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 9.74 (s, 1H, CO**H**), 3.85 (m, 1H, C**H**Cl), 2.43 (t, 2H, *J* = 7.2 Hz, COC$\mathbf{H}_2$), 1.77-1.20 (m, 12H, C$\mathbf{H}_2$), 0.87 (t, 3H, *J* = 7.2 Hz, C$\mathbf{H}_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 202.40, 63.82, 43.76, 38.24, 38.20, 28.66, 26.06, 22.26, 21.58, 13.99.

To a solution of methyl(triphenylphosphoranylidine)acetate (360 mg, 1.08 mmol) in dichloromethane (1 mL) at 0 ºC, the solution of (*S*)-6-chlorodecanal (171 mg, 0.898 mmol) in dichloromethane (2 mL) was added dropwise over 5 min. The reaction mixture was warmed to room temperature and reacted overnight. The reaction mixture was concentrated *in vacuo* and the crude product was purified by flash chromatography, eluting with 0-10% ethyl acetate in hexanes to afford methyl (*S*)-8-chlorododec-2-enoate **5.118** as a colorless oil (181 mg, 0.73 mmol, 82%). TLC: R_f = 0.73 (silica gel, 4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{13}H_{24}ClO_2^+$ $[M+H]^+$, 247.1459; found, 247.1474. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 6.94 (dt, 1H, *J* = 15.6, 7.0 Hz, C**H**CH_2), 5.81 (dt, 1H, *J* = 15.6, 1.6 Hz, COC**H**), 3.86 (m, 1H, C**H**Cl), 3.71 (s, 3H, OC$\mathbf{H}_3$), 2.21 (q, 2H, *J* = 6.3 Hz, C**H**CH_2), 1.75-1.63 (m, 4H, CHClC$\mathbf{H}_2$), 1.60-1.22 (m, 8H, C$\mathbf{H}_2$), 0.89 (t, 3H, *J* = 7.2 Hz, C$\mathbf{H}_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 167.0, 149.1, 121.1, 63.9, 51.3, 38.2 (2), 32.0, 28.6, 27.6, 26.0, 22.2, 13.9.

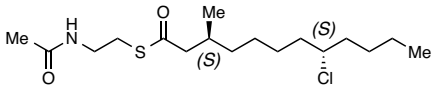


5.119

(3*S*,8*S*)-8-Chloro-3-methyldodecanoic acid (5.119): A suspension of copper iodide (2.3 mg, 0.01 mmol) and (*S*)-Tol-BINAP (12.4 mg, 0.02 mmol) in TBME (3 mL) was stirred at room temperature for 3 h until the solid turned yellow. The reaction mixture was cooled to –20 ºC and MeMgBr (3 M in ether, 1.01 mL, 3.04 mmol) was added dropwise over 5 min to the reaction mixture. After stirring for 15 min, a solution of (*S*)-methyl 8-chlorododec-2-enoate **5.118** (136 mg, 0.61 mmol) in TBME (4 mL) was added

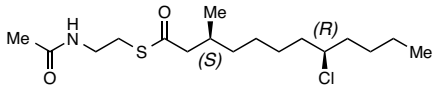
dropwise to the reaction mixture over 30 min. The reaction mixture was stirred at –25 ºC for 1 h and quenched with methanol (1 mL). The reaction mixture was diluted with a saturated aqueous solution of ammonium chloride (10 mL) and warmed to room temperature. The reaction mixture was extracted with ether (3 × 10 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 10% diethyl ether in hexanes to afford methyl (3*S*,8*S*)-8-chloro-3-methyldodecanoate, which was used directly in the next reaction without further purification. TLC: R_f = 0.64 (silica gel, 4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_{28}ClO_2^+$ $[M+H]^+$, 263.1772; found, 263.1779. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.87 (m, 1H, C**H**Cl), 3.66 (s, 3H, OC**H**$_3$), 2.29 (dd, 1H, *J* = 14.8, 6.1 Hz, COC**H**$_2$), 2.11 (dd, 1H, *J* = 14.8, 8.0 Hz, COC**H**$_2$), 1.95 (m, 1H, C**H**CH$_3$), 1.74-1.65 (m, 4H, CHClC**H**$_2$), 1.53-1.47 (m, 2H, C**H**$_2$), 1.42-1.18 (m, 8H, C**H**$_2$), 0.92 (m, 6H, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 173.8, 64.3, 51.5, 41.7, 38.6, 38.4, 36.7, 30.4, 28.8, 26.8, 26.6, 22.4, 19.9, 14.1.

Methyl (3*S*,8*S*)-8-chloro-3-methyldodecanoate (0.61 mmol) was dissolved in 2:2:1 solution of THF/water/ethanol (10 mL). Lithium hydroxide (26.5 mg, 1.11 mmol) was added to the reaction mixture, and the reaction mixture was stirred overnight at room temperature. The reaction mixture was acidified with 1 M HCl (15 mL) and extracted with dichloromethane (3 × 15 mL). The combined organic layers were washed with brine (20 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting at 50% ethyl acetate in hexanes to afford (3*S*,8*S*)-8-chloro-3-methyldodecanoic acid **5.119** as a colorless oil (86 mg, 0.35 mmol, 57%). TLC: R_f = 0.17 (silica gel, 3:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{13}H_{24}ClO_2^-$ $[M–H]^-$, 247.1470; found, 247.1465. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 11.69 (s, 1H, COO**H**), 3.88 (m, 1H, C**H**Cl), 2.35 (dd, 1H, *J* = 15.0, 6.0 Hz, COC**H**$_2$), 2.16 (dd, 1 H, *J* = 15.0, 8.1 Hz, COC**H**$_2$), 1.97 (m, 1H, C**H**CH$_3$), 1.78-1.64 (m, 4H, CHClC**H**$_2$), 1.58-1.46 (m, 2H, C**H**$_2$), 1.46-1.19 (m, 8H, C**H**$_2$), 0.97 (d, 3H, *J* = 6.7 Hz, CHC**H**$_3$), 0.91 (t, 3H, *J* = 7.2 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 179.8, 64.2, 41.5, 38.4, 38.2, 36.4, 30.1, 28.7, 26.6, 26.5, 22.3, 19.7, 14.0.



5.59

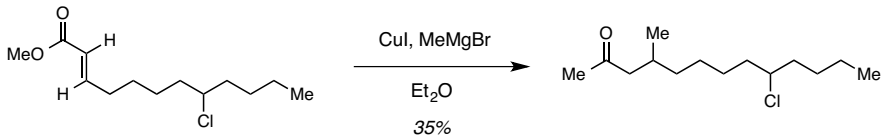
The (*S,S*)-chlorinated acyl-SNAC substrate **5.59** (7.7 mg, 0.022 mmol, 39%) was prepared from (3*S*,8*S*)-8-chloro-3-methyldodecanoic acid **5.119** (14 mg, 0.056 mmol) following the procedure used for the synthesis of **4.77** described in **Chapter 4.4.19**. TLC: R_f = 0.24 (silica gel, 1:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{17}H_{33}ClNO_2S^+$ $[M+H]^+$, 350.1915; found 350.1948: calc'd for $C_{17}H_{32}ClNNaO_2S^+$ $[M+Na]^+$, 372.1734; found 372.1758: calc'd for $C_{17}H_{32}ClNKO_2S^+$ $[M+K]^+$, 388.1474; found 388.1480. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 5.81 (s, 1H, N**H**), 3.89 (m, 1H, C**H**Cl), 3.44 (q, 2H, *J* = 6.2 Hz, NHC**H**$_2$), 3.04 (t, 3H, *J* = 6.4 Hz, CH_2C**H**$_2$S), 2.57 (dd, 1H, *J* = 14.6, 6.0 Hz, SCOC**H**$_2$), 2.40 (dd, 1H, *J* = 14.5, 8.1 Hz, SCOC**H**$_2$), 2.02 (m, 1H, C**H**), 1.97 (s, 3H, NHCOC**H**$_3$), 1.78-1.62 (m, 4H, C**H**$_2$CHCl), 1.57-1.46 (m, 2H, C**H**$_2$), 1.46-1.17 (m, 8H, C**H**$_2$), 0.94 (d, 3H, *J* = 6.7 Hz, C**H**$_3$), 0.92 (t, 3H, *J* = 7.3 Hz, C**H**$_3$). ^{13}C NMR (125 MHz; $CDCl_3$): δ 199. 9, 164.3, 64.4, 51.4, 40.0, 38.6, 38.4, 36.6, 31.2, 28.8, 28.7, 26.7, 26.6, 23.4, 22.4, 19.7, 14.1.



5.60

The (3*S,*8*R*)-chlorinated acyl-SNAC substrate (**5.60**) was synthesized following the same procedure as the preparation of the (3*S,*8*S*)-chlorinated acyl-SNAC substrate (**5.59**), except using the (*S*)-6-hydroxydecyl pivalate (**5.120**) as the starting material. The ^{1}H and ^{13}C NMR spectra of the (3*S,*8*R*)-chlorinated acyl-SNAC substrate (**5.60**) matched those of the (3*S,*8*S*)-chlorinated acyl-SNAC substrate (**5.59**).

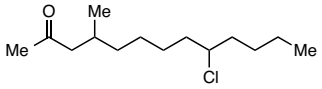
5.4.17. Synthesis of the substrates for screening CylK promiscuity



5.121 **5.66**

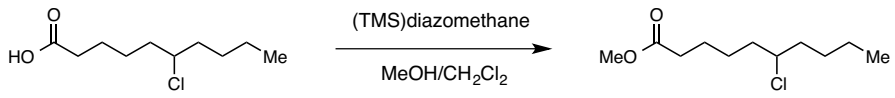
Scheme 5.27: Synthesis of the ketone substrate.

The synthesis of **5.121** is reported in **Chapter 4.4.17**.



5.66

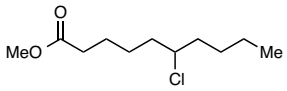
9-Chloro-4-methyltridecan-2-one (5.66): A suspension of copper iodide (120 mg, 0.63 mmol) in anhydrous diethyl ether (5 mL) was cooled to –20 ºC. To the suspension, methylmagnesium bromide (3 M in dietyl ether, 1.05 ml, 3.15 mmol) was added dropwise over 15 min, and the reaction mixture was stirred at –20 ºC for 1.5 h. Then, a solution of methyl 8-chlorododec-2-enoate **5.121** (154 mg, 0.63 mmol) in anhydrous ether (3 mL) was added dropwise over 15 min, and the reaction was stirred at –20 ºC for another 1.5 h. The reaction mixture was quenched by slow addition of methanol (1 mL) followed by saturated aqueous solution of sodium bicarbonate (30 mL). The reaction mixture was extracted with ethyl acetate (3 × 30 mL), and the combined organic fractions were washed with brine (40 mL), dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 10–15% diethyl ether in hexanes to afford 9-chloro-4-methyltridecan-2-one **5.66** (54 mg, 0.22 mmol, 35%). TLC: R_f = 0.70 (silica gel, 9:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{14}H_{27}ClNaO^+$ $[M+Na]^+$, 269.1643; found 269.1633: calc'd for $C_{14}H_{27}ClKO^+$ $[M+K]^+$, 285.1382; found 285.1404. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.87 (m, 1H, C**H**Cl), 2.39 (dd, 1H, *J* = 16.0, 5.7 Hz, COC**H**$_2$), 2.23 (dd, 1H, *J* = 15.9, 7.9 Hz, COC**H**$_2$), 2.12 (s, 3H, COC**H**$_3$), 1.98 (m, 1H, C**H**), 1.76-1.62 (m, 4H, CHClC**H**$_2$), 1.57-1.44 (m, 2H, C**H**$_2$), 1.44-1.11 (m, 8H, C**H**$_2$), 0.90 (m, 6H, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 209.1, 64.4, 51.4, 38.6, 38.4, 36.9, 30.6, 29.3, 28.8, 26.7, 26.7, 22.4, 19.9, 14.1.



5.68 **5.67**

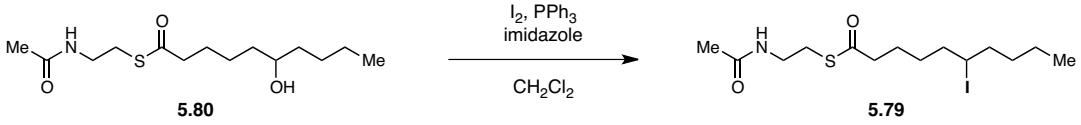
Scheme 5.28: Synthesis of the methyl ester.

The synthesis of **5.68** is described in **Chapter 4.4.17**.



5.67

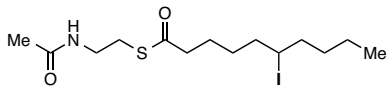
Methyl 6-chlorodecanoate (5.67): 6-Chlorodecanoic acid **5.68** (33 mg, 0.16 mmol) was dissolved in dichloromethane (1.2 mL) and methanol (0.8 mL). To the solution, (trimethylsilyl)diazomethane (2 M, 0.12 mL, 0.24 mmol) was added. The reaction mixture was stirred at room temperature for 30 min and then quenched with acetic acid (0.5 mL). The reaction mixture was diluted with ethyl acetate (10 mL), washed with saturated aqueous solution of sodium bicarbonate (2 × 10 mL) and brine (10 mL). The combined organic layers were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The crude product was purified by flash chromatography, eluting with 5% ethyl acetate in hexanes to afford methyl 6-chlorodecanoate **5.67** (28 mg, 0.13 mmol, 79%). TLC: R_f = 0.63 (silica gel, 4:1 hexanes/ethyl acetate). HRMS (ESI): calc'd for $C_{11}H_{22}ClO^+$ $[M+H]^+$, 221.1303; found 221.1323. ^{1}H-NMR (500 MHz; $CDCl_3$): δ 3.88 (m, 1H, C**H**Cl), 3.67 (s, 3H, OC**H**$_3$), 2.32 (t, 2H, *J* = 7.4 Hz, COC**H**$_2$), 1.78-1.23 (m, 12H, C**H**$_2$), 0.90 (t, 3H, *J* = 7.2 Hz, C**H**$_3$). ^{13}C-NMR (125 MHz; $CDCl_3$): δ 174. 1, 64.0, 51.7, 38.4, 38.3, 34.1, 28.8, 26.2, 24.6, 22.4.



5.80 **5.79**

Scheme 5.29: Synthesis of the 6-iododecanoyl-SNAC substrate.

The synthesis of **5.80** is described in **Chapter 2.4.16**.



5.79

6-iododecanoyl-SNAC (5.79): A 25 mL reaction flask was charged with SNAC derivative **5.80** (70 mg, 0.24 mmol), triphenylphosphine (140 mg, 0.53 mmol) and dichloromethane (5 mL). The resulting mixture was cooled to 0 °C and imidazole (36 mg, 0.53 mmol) then iodine (150 mg, 0.60 mmol) were added to the reaction flask. The reaction mixture was warmed to room temperature and stirred overnight. The reaction mixture was then cooled to 0 °C and quenched with the addition of saturated aqueous sodium thiosulfate (5 mL), and transferred into a separatory funnel. The phases were separated and the aqueous layer was extracted with dichloromethane (3 x 5 mL). The combined organic layers were dried over sodium sulfate, filtered and concentrated *in vacuo*. The crude residue was purified by column chromatography eluting with ethyl acetate to give the 6-iododecanoyl-SNAC derivative **5.79** (23 mg, 0.058 mmol, 24%) as a colorless oil. TLC: R_f = 0.42 (ethyl acetate). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.87 – 5.78 (bs, 1H, **NH**), 4.13 – 4.05 (m, 1H), 3.44 (q, *J* = 6.2 Hz, 2H, NC**H**$_2$), 3.03 (t, *J* = 6.4 Hz, 2H, SC**H**$_2$), 2.60 (t, *J* = 7.4 Hz, 2H, SCOC**H**$_2$), 1.97 (s, 3H), 1.61 (m, 4H, C**H**$_2$), 1.52 – 1.21 (m, 8H, C**H**$_2$), 0.91 (t, *J* = 7.2 Hz, 3H, C**H**$_3$).

5.5. References

(1) Fessner, W. D. Enzyme mediated C–C bond formation. *Curr. Opin. Chem. Biol.* **1998**, *2*, 85.

(2) Fesko, K.; Gruber-Khadjawi, M. Biocatalytic methods for C–C bond formation. *ChemCatChem* **2013**, *5*, 1248.

(3) Dresen, C.; Richter, M.; Pohl, M.; Ludeke, S.; Müller, M. The enzymatic asymmetric conjugate umpolung reaction. *Angew. Chem. Int. Ed.* **2010**, *49*, 6600.

(4) Svedendahl, M.; Hult, K.; Berglund, P. Fast carbon–carbon bond formation by a promiscuous lipase. *J. Am. Chem. Soc.* **2005**, *127*, 17988.

(5) He, Y.-H.; He, T.; Guo, J.-T.; Li, R.; Xiang, Y.; Yang, D.-C.; Guan, Z. Enzyme-catalyzed domino reaction: Efficient construction of spirocyclic oxindole skeleton using porcine pepsin. *Catal. Sci. Technol.* **2016**.

(6) Ogura, K.; Koyama, T. Enzymatic aspects of isoprenoid chain elongation. *Chem. Rev.* **1998**, *98*, 1263.

(7) Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **2006**, *106*, 3412.

(8) Cane, D. E. Enzymatic formation of sesquiterpenes. *Chem. Rev.* **1990**, *90*, 1089.

(9) Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. *S*-adenosylmethionine-dependent methyltransferases: Highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* **2012**, *13*, 2642.

(10) Choi, J. M.; Han, S. S.; Kim, H. S. Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnol. Adv.* **2015**, *33*, 1443.

(11) Porter, J. L.; Rusli, R. A.; Ollis, D. L. Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem* **2016**, *17*, 197.

(12) Resch, V.; Schrittwieser, J. H.; Siirola, E.; Kroutil, W. Novel carbon–carbon bond formations for biocatalysis. *Curr. Opin. Biotechnol.* **2011**, *22*, 793.

(13) Chen, J. L.; Moore, R. E.; Patterson, G. M. L. Structures of nostocyclophanes A-D. *J. Org. Chem.* **1991**, *56*, 4360.

(14) Nakamura, H.; Hamer, H. A.; Sirasani, G.; Balskus, E. P. Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J. Am. Chem. Soc.* **2012**, *134*, 18518.

(15) Preisitsch, M.; Niedermeyer, T. H. J.; Heiden, S. E.; Neidhardt, I.; Kumpfmuller, J.; Wurster, M.; Harmrolfs, K.; Wiesner, C.; Enke, H.; Müller, R.; Mundt, S. Cylindrofridins A-C, linear cylindrocyclophane-related alkylresorcinols from the cyanobacterium *Cylindrospermum stagnale*. *J. Nat. Prod.* **2016**, *79*, 106.

(16) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M.; Orjala, J. Cylindrocyclophanes with proteasome inhibitory activity from the cyanobacterium *Nostoc* sp. *J. Nat. Prod.* **2010**, *73*, 1529.

(17) Preisitsch, M.; Heiden, S. E.; Beerbaum, M.; Niedermeyer, T. H.; Schneefeld, M.; Herrmann, J.; Kumpfmuller, J.; Thurmer, A.; Neidhardt, I.; Wiesner, C.; Daniel, R.; Müller, R.; Bange, F. C.; Schmieder, P.; Schweder, T.; Mundt, S. Effects of halide ions on the carbamidocyclophane biosynthesis in *Nostoc* sp. CAVN2. *Mar. Drugs* **2016**, *14*.

(18) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E. Structures of cylindrocyclophanes A-F. *Tetrahedron* **1992**, *48*, 3001.

(19) Preisitsch, M.; Harmrolfs, K.; Pham, H. T. L.; Heiden, S. E.; Fuessel, A.; Wiesner, C.; Pretsch, A.; Swiatecka-Hagenbruch, M.; Niedermeyer, T. H. J.; Müller, R.; Mundt, S. Anti-MRSA-acting carbamidocyclophanes H-L from the Vietnamese cyanobacterium *Nostoc* sp. CAVN2. *J. Antibiot.* **2015**, *68*, 600.

(20) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567.

(21) Pasquier, C.; Promponas, V. J.; Palaios, G. A.; Hamodrakas, J. S.; Hamodrakas, S. J. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the Swiss-Prot database: The PRED-TMR algorithm. *Protein Eng.* **1999**, *12*, 381.

(22) Soding, J.; Biegert, A.; Lupas, A. N. The HHPred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244.

(23) Yang, J.; Kulkarni, K.; Manolaridis, I.; Zhang, Z. G.; Dodd, R. B.; Mas-Droux, C.; Barford, D. Mechanism of isoprenylcysteine carboxyl methylation from the crystal structure of the integral membrane methyltransferase icmt. *Mol. Cell* **2011**, *44*, 997.

(24) Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in *Bacillus subtilis*. *J. Bacteriol.* **2009**, *191*, 4916.

(25) Nakano, C.; Funa, N.; Ohnishi, Y.; Horinouchi, S. The *O*-methyltransferase SrsB catalyzes the decarboxylative methylation of alkylresorcylic acid during phenolic lipid biosynthesis by *Streptomyces griseus*. *J. Bacteriol.* **2012**, *194*, 1544.

(26) Funabashi, M.; Funa, N.; Horinouchi, S. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on *Streptomyces griseus*. *J. Biol. Chem.* **2008**, *283*, 25104.

(27) Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S. N.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z. X.; Yamashita, R. A.; Zhang, D. C.; Zheng, C. J.; Bryant, S. H. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222.

(28) Linhartova, I.; Bumba, L.; Masin, J.; Basler, M.; Osicka, R.; Kamanova, J.; Prochazkova, K.; Adkins, I.; Hejnova-Holubova, J.; Sadilkova, L.; Morova, J.; Sebo, P. RTX proteins: A highly diverse family secreted by a common mechanism. *FEMS Microbiol. Rev.* **2010**, *34*, 1076.

(29) Goebel, W.; Hedgpeth, J. Cloning and functional characterization of the plasmid encoded hemolysin determinant of *Escherichia coli*. *J. Bacteriol.* **1982**, *151*, 1290.

(30) Muller, D.; Hughes, C.; Goebel, W. Relationship between plasmid and chromosomal hemolysin determinants of *Escherichia coli*. *J. Bacteriol.* **1983**, *153*, 846.

(31) Felmlee, T.; Welch, R. A. Alterations of amino acid repeats in the *Escherichia coli* hemolysin affect cytolytic activity and secretion. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5269.

(32) Slack, F. J.; Ruvkun, G. A novel repeat domain that is often associated with ring finger and B-box motifs. *Trends Biochem. Sci.* **1998**, *23*, 474.

(33) Chen, C. K. M.; Chan, N. L.; Wang, A. H. J. The many blades of the beta-propeller proteins: Conserved but versatile. *Trends Biochem. Sci.* **2011**, *36*, 553.

(34) Fulop, V.; Jones, D. T. Beta propellers: Structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* **1999**, *9*, 715.

(35) Buchinger, E.; Knudsen, D. H.; Behrens, M. A.; Pedersen, J. S.; Aarstad, O. A.; Tondervik, A.; Valla, S.; Skjak-Braek, G.; Winnmer, R.; Aachmann, F. L. Structural and functional characterization of

the R-modules in alginate C5 epimerases AlgE4 and AlgE6 from *Azotobacter vinelandii. J. Biol. Chem.* **2014**, *289*, 31382.

(36) Korczynska, M.; Mukhtar, T. A.; Wright, G. D.; Berghuis, A. M. Structural basis for streptogramin B resistance in *Staphylococcus aureus* by virginiamycin B lyase. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10388.

(37) Ertesvag, H.; Hoidal, H. K.; Hals, I. K.; Rian, A.; Doseth, B.; Valla, S. A family of modular type mannuronan C5 epimerase genes controls alginate structure in *Azotobacter vinelandii*. *Mol. Microbiol.* **1995**, *16*, 719.

(38) Franklin, M. J.; Chitnis, C. E.; Gacesa, P.; Sonesson, A.; White, D. C.; Ohman, D. E. *Pseudomonas aeruginosa* AlgG is a polymer level alginate C5-mannuronan epimerase. *J. Bacteriol.* **1994**, *176*, 1821.

(39) Skjakbraek, G.; Larsen, B. Biosynthesis of alginate—purification and characterization of mannuronan C5 epimerase from *Azotobacter vinelandii. Carbohydr. Res.* **1985**, *139*, 273.

(40) Hoidal, H. K.; Ertesvag, H.; Skjak-Braek, G.; Stokke, B. T.; Valla, S. The recombinant *Azotobacter vinelandii* mannuronan C5-epimerase AlgE4 epimerizes alginate by a nonrandom attack mechanism. *J. Biol. Chem.* **1999**, *274*, 12316.

(41) Campa, C.; Holtan, S.; Nilsen, N.; Bjerkan, T. M.; Stokke, B. T.; Skjak-braek, G. Biochemical analysis of the processive mechanism for epimerization of alginate by mannuronan C5 epimerase AlgE4. *Biochem. J.* **2004**, *381*, 155.

(42) Mukhtar, T. A.; Koteva, K. P.; Hughes, D. W.; Wright, G. D. Vgb from staphylococcus aureus inactivates streptogramin B antibiotics by an elimination mechanism not hydrolysis. *Biochemistry* **2001**, *40*, 8877.

(43) Kang, H. S.; Santarsiero, B. D.; Kim, H.; Krunic, A.; Shen, Q.; Swanson, S. M.; Chai, H.; Kinghorn, A. D.; Orjala, J. Merocyclophanes A and B, antiproliferative cyclophanes from the cultured terrestrial cyanobacterium *Nostoc* sp. *Phytochemistry* **2012**, *79*, 109.

(44) Herron, S. R.; Benen, J. A. E.; Scavetta, R. D.; Visser, J.; Jurnak, F. Structure and function of pectic enzymes: Virulence factors of plant pathogens. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8762.

(45) Seyedarabi, A.; To, T. T.; Ali, S.; Hussain, S.; Fries, M.; Madsen, R.; Clausen, M. H.; Teixteira, S.; Brocklehurst, K.; Pickersgill, R. W. Structural insights into substrate specificity and the anti beta-elimination mechanism of pectate lyase. *Biochemistry* **2010**, *49*, 539.

(46) Leão, P. N.; Nakamura, H.; Costa, M.; Pereira, A. R.; Martins, R.; Vasconcelos, V.; Gerwick, W. H.; Balskus, E. P. Biosynthesis-assisted structural elucidation of the bartolosides, chlorinated aromatic glycolipids from cyanobacteria. *Angew. Chem. Int. Ed.* **2015**, *54*, 11063.

(47) Olah, G. A. *Friedel-crafts chemistry*; Wiley: New York,, 1973.

(48) Heide, L. Prenyl transfer to aromatic substrates: Genetics and enzymology. *Curr. Opin. Chem. Biol.* **2009**, *13*, 171.

(49) Hammer, S. C.; Dominicus, J. M.; Syren, P. O.; Nestl, B. M.; Hauer, B. Stereoselective Friedel–Crafts alkylation catalyzed by squalene hopene cyclases. *Tetrahedron* **2012**, *68*, 7624.

(50) Hammer, S. C.; Syren, P. O.; Seitz, M.; Nestl, B. M.; Hauer, B. Squalene hopene cyclases: Highly promiscuous and evolvable catalysts for stereoselective C–C and C–X bond formation. *Curr. Opin. Chem. Biol.* **2013**, *17*, 293.

(51) Stecher, H.; Tengg, M.; Ueberbacher, B. J.; Remler, P.; Schwab, H.; Griengl, H.; Gruber-Khadjawi, M. Biocatalytic Friedel–Crafts alkylation using non-natural cofactors. *Angew. Chem. Int. Ed.* **2009**, *48*, 9546.

(52) Tengg, M.; Stecher, H.; Offner, L.; Plasch, K.; Anderl, F.; Weber, H.; Schwab, H.; Gruber-Khadjawi, M. Methyltransferases: Green catalysts for Friedel–Crafts alkylations. *ChemCatChem* **2016**, *8*, 1.

(53) Zhou, K.; Ludwig, L.; Li, S. M. Friedel–Crafts alkylation of acylphloroglucinols catalyzed by a fungal indole prenyltransferase. *J. Nat. Prod.* **2015**, *78*, 929.

(54) Fan, A.; Xie, X.; Li, S. M. Tryptophan prenyltransferases showing higher catalytic activities for Friedel–Crafts alkylation of *o*- and *m*-tyrosines than tyrosine prenyltransferases. *Org. Biomol. Chem.* **2015**, *13*, 7551.

(55) Mori, T.; Zhang, L.; Awakawa, T.; Hoshino, S.; Okada, M.; Morita, H.; Abe, I. Manipulation of prenylation reactions by structure-based engineering of bacterial indolactam prenyltransferases. *Nat. Commun.* **2016**, *7*, 10849.

(56) Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Bio.* **2009**, *10*, 866.