

Investigation of atypical transformations during the biosynthesis of marine
cyanobacterial natural products

by

Eli Benchell Eisman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2013

Doctoral Committee:

Professor David H. Sherman, Chair
Associate Professor Kristina I. Hakansson
Professor Anna K. Mapp
Professor Janet L. Smith

*This work is dedicated to the memory of Dr. Tomoko Komiyama.
An exemplary scientist and honest human being.*

Acknowledgments

Without question, the person most responsible for my degree completion is my advisor Professor David H. Sherman. David welcomed me into his group and through many, very challenging years, I have reached a goal that I never thought achievable. David pushed me beyond my limit in the best way possible and I am incredibly thankful for his limitless patience and support. Thank you.

My thesis committee, composed of Professors Kristina Hakansson, Anna K. Mapp, and Janet L. Smith, offered guidance, advice and encouragement during the course of my tenure as a graduate student, for which I am exceedingly grateful. I do not believe that I would have been as successful without their collective assistance. Thank you.

I have been fortunate enough to work with many excellent collaborators over the last five years, but I would like to highlight Dr. Jennifer Gehret-McCarthy for her contribution to my overall development as a researcher. Thank you. Additionally, Chris Kubitskey, a motivated and self-directed undergraduate student, enabled much of the work in the last six months of degree timeline. I am sure that he will be very successful in his future goals, whatever they may be. Thank you.

Lastly, my family, friends, colleagues have all empowered me to complete this task, especially Karoline Chiou and Kyle Bolduc who have welcomed me into their life. Shamilya Therese Williams is the backbone of the Sherman group and I am very thankful have had her assistance in navigating through graduate school.

I have done many things with my life so far and for the first time in a very long time I am proud of what I have accomplished. I love you all. Dynamic and powerful.

Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
List of figures.....	vii
List of tables.....	x

Chapter 1-Natural products are bioactive small molecules from the environment.....	1
1.1 Natural products have been in use as medicines for millennia.....	1
1.2 Bioactive secondary metabolites are produced by enzyme megasynthases.....	1
1.3 Moorea producens natural products.....	5
1.4 Curacin biosynthesis.....	7
1.4.1 β -branching cassette leads to 2-methyl-cyclopropyl ring formation.....	9
1.4.2 Activation-elimination offloading strategy leading to olefin formation.....	11
1.5 Research outline.....	13

Chapter 2- Noncatalytic C-terminal protein domain enhances turnover during curacin β-branching.....	14
2.1 Introduction.....	14
2.2 Results.....	16
2.2.1 Cd increases biosynthetic flux of a reconstituted β -branching system.....	17
2.2.2 Biophysical characterization of Cd on protein dynamics.....	18
2.2.3 Preliminary characterization of Cd catalyzed self-association.....	21
2.2.4 Mechanistic insights into cyclopropane formation by individual reaction monitoring.....	22
2.2.4.1 Cd dramatically increases rate of halogenation.....	24
2.2.4.2 Cd does not affect rate, but increases extent of dehydration.....	25
2.2.4.3 Cd does not effect rate, but increases extent of decarboxylation.....	26
2.2.4.4 ECH2 N-terminus has minimal effect on rate of decarboxylation.....	26
2.2.4 Mechanistic insights into cyclopropane formation by individual reaction monitoring.....	22
2.2.4.1 Cd dramatically increases rate of halogenation.....	24
2.2.4.2 Cd does not affect rate, but increases extent of dehydration.....	25

2.2.4.3 Cd does not effect rate, but increases extent of decarboxylation.....	26
2.2.4.4 ECH2 N-terminus has minimal effect on rate of decarboxylation	26
2.2.4.5 Cd does not effect rate, but increases extent of cyclization	27
2.2.4.6 Reconstitution of β -branching cassette in a one-pot reaction	28
2.2.5 Solution structure of ACPI and characterization of Hal-ACP interaction by MS	30
2.2.6 Interrogation of Hal-ACP interaction by computational docking.....	33
2.2.7 Putative structure of the Cd resembles PKS docking domains	36
2.3 Discussion	37
2.4 Materials and methods	39
Chapter 3- Activated-elimination as a strategy for hydrocarbon production	46
3.1 Introduction	46
3.2 Results-Sulfotransferase-Thioesterase containing genes from additional cyanobacteria	47
3.2.1 SynTE is capable of accepting Cur substrate but does not yield alkene.....	48
3.2.2 SynTE HRSC double mutant prepared for active site labeling protein	52
3.2.3 Alternative methods of olefin detection via alkene functionalization	54
3.2.4 Syn and Cur proteins act functionally in trans and accept nonnative substrates	55
3.2.5 ST structure and loop exchange	60
3.2.6 Chemical and enzymatic access of acyl-CoA substrates	61
3.2.6.1 Acyl activating enzymes couple free fatty acids to holo-carrier proteins.....	63
3.3 Discussion.....	67
3.4 Materials and methods	69
Chapter 4- Origins of the linear lipopeptide carmabin A	75
4.1 Introduction	75
4.2 Carmabin pathway	77
4.3 Proposed carmabin A biosynthesis	80
4.4 Probing alkyne formation	81
4.4.1 Overexpression of CarB in a eukaryotic host	82
4.4.2 Fatty acid method development.....	83
4.4.2.1 Gas chromatography is a standard method for hydrocarbon analysis	83
4.4.2.2 Headspace analysis of volatile compounds.....	86
4.5 Carmabin A possesses a terminal amide.....	87

4.5.1 CarI Terminal adenylation domain	87
4.5.2 CarI adenylation-methyltransferase-methyltransferase tridomain.....	90
4.6 Discussion.....	92
4.7 Materials and methods	95
Chapter 5- Future directions	99
5.1 Megasyntase enzymology.....	99
5.1.1 Type-I PKS modular strategy	100
5.1.2 Iterative Type-I PKS enzymology	101
5.1.3 Acyl carrier proteins as active participants during small molecule assembly	102
5.2 Chain elongation intermediates as pseudo-activity based protein-profiling probes	102
References	105

List of figures

Figure 1. Polyketide synthase (PKS) enzyme strategy	2
Figure 2. Nonribosomal peptide synthetase (NRPS) enzyme strategy	3
Figure 3. PKS reductive enzymes increase structural diversity in natural products.....	4
Figure 4. NRPS tailoring domains increase structural diversity in natural products.....	4
Figure 5. Natural products isolated from the marine cyanobacterium <i>Moorea producens</i> ..	5
Figure 6. Curacin A biosynthetic pathway.	8
Figure 7. Curacin A β -branching assembly pathway.....	10
Figure 8. Curacin A olefinative offloading.....	12
Figure 9. Natural products harboring β -branching functionalization. the protein cassette.....	15
Figure 10. Curacin A 2-methyl-cyclopropane is formed by β -branching protein	16
Figure 11. Sequence alignment of excised ACP domains from the CurA and JamE carrier protein multiplets.....	16
Figure 12. Sequence alignment of the terminal C-terminal flanking domains, Cd, from the CurA and JamE carrier protein multiplets..	17
Figure 13. Schematic for one-pot reaction and metabolite analysis..	17
Figure 14. Analytical size exclusion chromatography for CurA ACP multiplets	18
Figure 15. Electron microscopy analysis of CurA ACP multiplets by negative staining.....	21
Figure 16. ITC dilution assay for the CurA Cd.	22
Figure 17. Examination of discrete β -branching interactions by LC-MS.....	24
Figure 18. Cd dramatically increases rate of halogenation of HMG-ACP.....	24
Figure 19. Effect of Cd on ECH1 activity	25
Figure 20. Evaluation of ACP isoform following ECH1 interaction.....	26
Figure 21. Effect of Cd on ECH2 activity	26
Figure 22. ECH2 activity is independent of N-terminal docking domain and presence of Cd	27
Figure 23. Effect of Cd on ER activity	27
Figure 24. Effect of Cd upon β -branching reconstitution.....	28
Figure 25. Distribution ACP isoform over time in the one-pot assay cyclopropanation assay, ACPIII, +9 charge state	29

Figure 26. Distribution ACP isoform over time in the one-pot assay cyclopropanation assay, ACP ^{III} +Cd ⁺ , +11 charge state.....	29
Figure 27. Role assignment for individual β -branching reactions during cyclopropane formation based on LC-MS analysis	30
Figure 28. Analysis of ACP-Hal interaction via LC-MS/MS.....	32
Figure 29. Computationally docked model for CurA Hal and ACP.....	33
Figure 30. ACP-Hal active site interaction.....	34
Figure 31. Salt bridge interactions are disturbed upon ACP D1988 mutation.	34
Figure 32. ACP hydrogen bonding is disrupted in site directed mutants.	35
Figure 33. Tightly packed hydrophobic network on the ACP is disrupted by mutation.	36
Figure 34. Structural comparison of donor modules from intermodular system.....	36
Figure 35. Offloading of curacin A desulfonative-decarboxylation.....	47
Figure 36. Standalone cyanobacterial pathways identified to contain CurM-like ST-TE domain	48
Figure 37. Standards used for GC-MS method development for hydrocarbon detection	49
Figure 38. Schematic for sulfotransferase and thioesterase catalyzed reactions	50
Figure 39. GC-MS analysis of ST-TE reaction workups.....	50
Figure 40. LC-MS analysis of the product of CurST and SynTE.....	51
Figure 41. Structural comparison of hydrolysing and decarboxylating thioesterases	52
Figure 42. LC-MS analysis of SynTE HRSC labeling.....	53
Figure 43. Thioclick enrichment of alkene containing compounds.....	55
Figure 44. LC-MS analysis of the product of CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA	57
Figure 45. LC-MS analysis of the product of CurST and CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA	58
Figure 46 LC-MS analysis of the product of CurST, CurTE CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA.	59
Figure 47. LC-MS analysis of the product of SynST and CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA	60
Figure 48. LC-MS analysis of FadD6 catalyzed formation of 3-OH-tetradecanoyl-CoA.....	62
Figure 49. Alignment of fatty acyl CoA ligases and fatty acyl AMP ligases.....	63
Figure 50. Reaction scheme for examining AAE activity	64

Figure 51. Alignment of ACPs from CurM and Syn olefin synthase.....	64
Figure 52. TIC shows formation of new peak upon incubation of Syn AAE with Syn ACP1.....	65
Figure 53. Deconvoluted mass abundances for AAE loaded SynLDG-ACP.....	66
Figure 54. Zoom of average mass spectrum from 3.330-3.586 min for AAE loaded SynLDG-ACP.....	67
Figure 55. Cyanobacterial natural products containing an alkyne and amide	76
Figure 56. Proposed carmabin A biosynthetic pathway	79
Figure 57. Alignment of bacterial and plant fatty acid desaturases	81
Figure 58. Density alignment score of fatty acid desaturases from carmabin and jamaicamide pathways	82
Figure 59. Microsome purification of CarB from yeast	83
Figure 60. Short chain fatty acids separation by GC	84
Figure 61. Fatty acids incubated in buffer show reduced signal.....	85
Figure 62. Cosolvents do not increase fatty acid detection by GC.....	85
Figure 63. HPLC-coupled evaporative light scattering detection enables separation and visualization of short chain fatty acids	87
Figure 64. Cladogram for putative functions of the terminal domain from CarI	88
Figure 65. Pyrophosphate exchange assay to determine CarI terminal adenylation domain amino acid specificity. Radioactive	89
Figure 66. Overexpression of CarI AMM is dramatically increased with coexpression of protein chaperone	91
Figure 67. Overexpression of CarI AMM is slightly increased upon inclusion of CHAPSO detergent.....	91
Figure 68. Overexpression of CarI AMM is slightly increased upon inclusion of chemical chaperone	92

List of tables

Table 1. Natural and synthetic analogs of curacin retain bioactivity.....	7
Table 2. Apparent molecular weights of ACPs by analytical gel filtration chromatography.....	19
Table 3. Biophysical constants obtained for CurA carrier proteins by analytical ultracentrifugation.....	20
Table 4. Deduced functions of the Open Reading Frames in the car gene cluster. Domain predictions characterized by protein BLAST assignment.	80
Table 5. Carmabin adenylation domain specificities profiles.....	80

Chapter 1

Natural products are bioactive small molecules from the environment

1.1 Natural products have been in use as medicines for millennia

During prehistory herbalists and shamans were known to utilize natural products (NP) in order to alleviate medical ailments. Often in the form of tinctures, salves and elixirs, early physicians made use of what crude natural samples were available at the time in the formation of the earliest pharmaceuticals (*I*). One of the best examples of harnessing the power of the environment for medical benefit is that of the *Salix alba* plant. First discussed by Greek physician Hippocrates in the fifth century BCE, the dried, crystalline powder resulting from resin dried willow bark resin was noted for its ability to ease aches, relax pains and reduce fevers. In 1826, German chemist Johann Buchner isolated the active compound, which he termed salicin. By stripping the sugar and oxidizing the compound to the benzoic acid, Italian chemist Raffaele Piria, was able to synthesize salicylic acid, the chemical precursor for the aspirin. Since then pharmacognosy has become increasingly important to the modern pharmaceutical paradigm (*2-5*). As of 2012, approximately 80% of all drugs available on the U.S. market were themselves NPs, NP-derivatives or NP-inspired chemicals (*6, 7*).

1.2 Bioactive secondary metabolites are produced by enzyme megasynthases

The observed breadth of biological activity of these environmentally isolable small molecules is predicated on a complex molecular architecture. The vast sampling of chemical space is built on a scaffold of subunits assembled by a unique set of coevolved modular enzymatic machinery (*8-II*). Polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) are large

multidomain enzymes that catalyze the stepwise addition of simple chemicals in a prototypically unidirectional fashion, in order to generate small molecules with potent biological activity.

PKSs and NRPSs share both a common strategy and domain architecture in the biosynthesis of NPs. The three core PKS domains are found in the order, ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP). The ACP is a noncatalytic protein that is posttranslationally modified at an invariant serine with a Coenzyme A (CoA) derived phosphopantetheine (PPant), the terminal thiol of which acts as a chemical tether responsible for shuttling the nascent chain between catalytic domains. AT domains selectively transfer simple acyl-Coenzyme A derivatives, such as malonic acid, onto the PPant arm of the downstream ACP domain. KS domains are bifunctional catalyzing the transfer of the nascent intermediate from the upstream or donor module to a cysteine in the KS active site of the acceptor module. Subsequent decarboxylation of the elongation unit appended to the downstream ACP and Claisen-like condensation with the growing chain, constituting one round of canonical PKS elongation (Figure 1) (8).

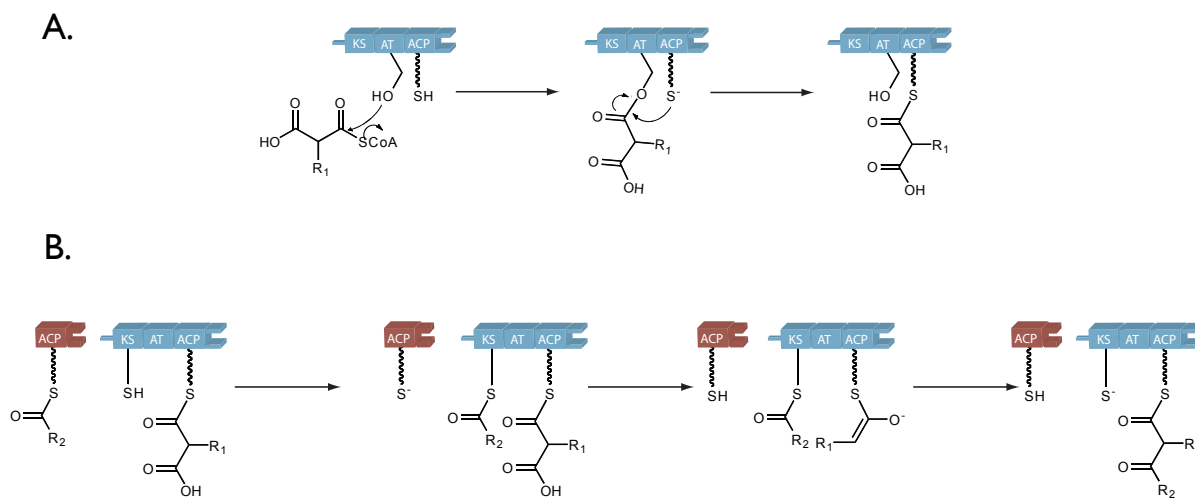


Figure 1. Polyketide synthase (PKS) enzyme strategy A. Acyltransferase (AT) catalyzed activation of short chain acyl-CoAs and transfer to acyl carrier protein (ACP) B. Intermodule interaction and ketosynthase (KS) catalyzed bond formation.

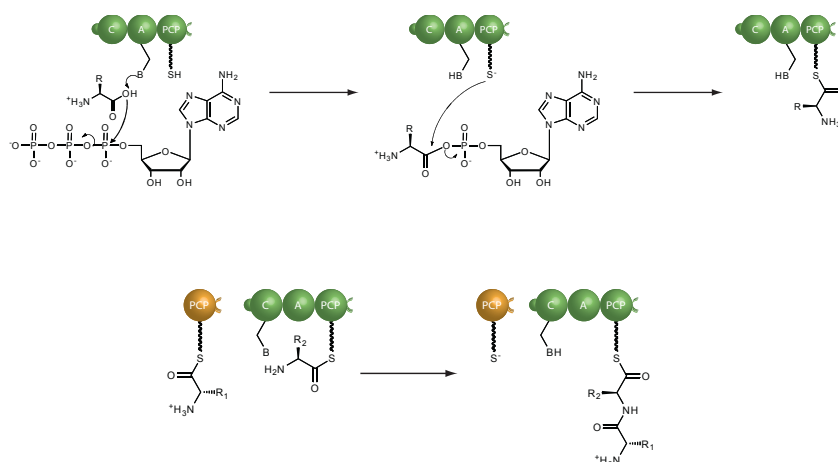


Figure 2. Nonribosomal peptide synthetase (NRPS) enzyme strategy A. Adenylation (A) domain catalyzed activation of short amino acids as amino acyl-AMP and transfer to peptidyl carrier protein (PCP) B. Intermodule interaction and condensation (C) domain catalyzed peptide bond formation.

Similarly, NRPS modules have the domain order condensation (C), adenylation (A), peptidyl carrier protein (PCP), which serves the homologous function as ACPs (Figure 2). A domains, recognize and activate amino acids as aminoacyl-AMP intermediates which are then thioesterified to the PPant arm of the downstream PCP domain. The C domain catalyzes base abstraction of a hydrogen the upstream amide group and subsequent nucleophilic attack by the NH_3 group of the PCP bound amino acid and chain elongation by the formation of a new peptide bond (**10**).

Diversity in natural products is rapidly and easily accessed through atypical substrate incorporation during metabolite construction. For PKS assemblies, this includes β -keto acids with variable substituents at the α -position, while for NRPSs the use of proteinogenic and nonproteinogenic amino acids, α -keto acids and other unique monomers account for the observed bricolage of NPs. Structural heterogeneity is also achieved by elaborating the core biosynthetic modules through inclusion of in-line tailoring enzymes. In PKSs these can include ketoreductases (KR), which catalyze stereospecific reduction of the β -keto group to either the (*R*) or (*S*) alcohol, dehydratases (DH) which, using the KR generated alcohol, form an α - β unsaturation in the growing chain and enoylreductase (ER) which reduce the DH product to the fully saturated alkane subunit (Figure 3).

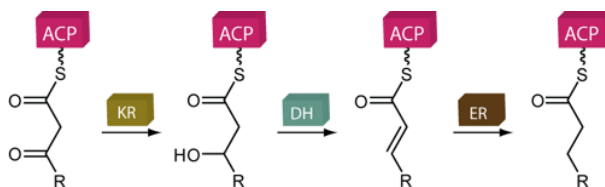


Figure 3. PKS reductive enzymes increase structural diversity in natural products.

Auxiliary NRPS domains include epimerases (E), which convert the stereochemistry of the added amino acid (i.e. from D- to L-), methyltransferases (MT) which are responsible for addition of a methyl group to the growing chain, typically at the amide nitrogen, or distal heteroatom of the amino acid R-group and cyclases (Cy) which cyclize medium and long chain amino acids such as cysteine into thiazoline (**12**) moieties (Figure 4). Additionally, the incorporation of both PKS and NRPS proteins within a single biosynthetic pathway, giving rise to a mixed system serves to further increase the number of small molecules that can be produced.

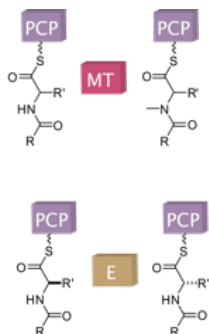


Figure 4. NRPS tailoring domains increase structural diversity in natural products.

The mechanism of termination is also a means by which structural variability is introduced during NP biosynthesis (**13**). Off-loading of the growing chain is typically achieved by the functioning of a domain found at the C terminus of the final protein module. Such enzymes include thioesterases (TE) which function as hydrolytic enzymes yielding a free acid, or catalyzing intramolecular cyclization generating either a macrolactone, as in pikromycin (**14**), or a macrolactam, observed in syringolin biosynthesis (**15**). Reductases (Rd) can catalyze two-electron reduction to release the NP as an aldehyde, observed in assembly of ET-743 (**16**), or the Rd can catalyze a four-electron reduction yielding an alcohol, as in myxalamid off-loading (**17**).

1.3 *Moorea producens* natural products

Over the last two decades drug discovery efforts have begun to expand from terrestrial sources and into the aquatic environment (18, 19). So-called bioprospecting has yielded a high number of metabolites with incredible structural diversity (20). Largely, natural products have been isolated from both freshwater and marine invertebrates. Of particular focus are natural products derived from marine cyanobacteria, from which over 300 nitrogen containing secondary metabolites have been reported (21). Notable amongst blue-green algae is the filamentous cyanobacterium, *Moorea producens* (22), previously *Lyngbya majuscula* (23), as a prolific producer of bioactive metabolites. These include the molluscicidal barbamide (24), protumor agent lyngbyatoxin (25), antifungal hectochlorin (26), neurotoxic jamaicamides (27), antimalarial carmabins (28), and the anticancer compound curacin A (12) (Figure 5).

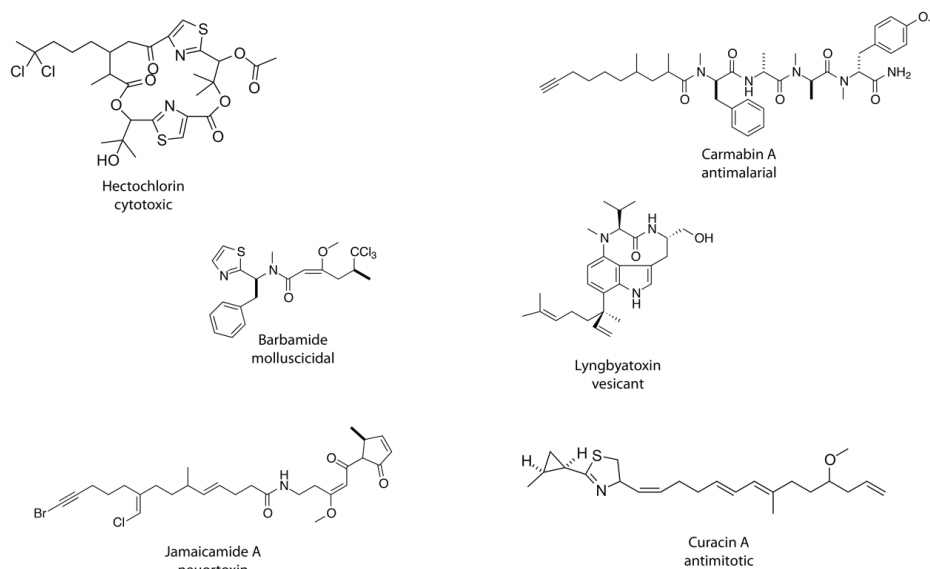


Figure 5. Natural products isolated from the marine cyanobacterium *Moorea producens*.

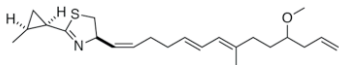
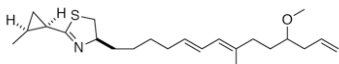
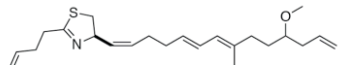
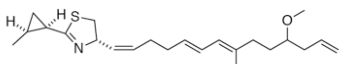
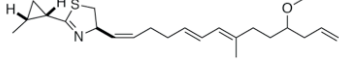
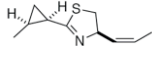
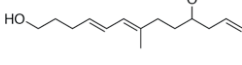
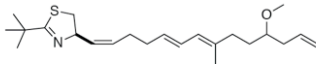
1.3.1 Curacin A is a potent antimitotic

First isolated by Gerwick and coworkers in 1994, by bioassay-guided fractionation, the linear lipopeptide curacin A was shown to be active against both leukemia and lymphoma cancer cell lines, with a remarkable 3-fold greater potency than the clinically used plant natural product, vincristine (29). As with other tubulin-targeting mitotic inhibitors (30), the cellular activity was determined to derive from G₂/M transition arrest (31). The inhibition of microtubule polymerization was determined to result from binding of the curacin A at the colchicine site on

tubulin (**31**). Moreover, Verdier-Pinard *et al.* (**32**) demonstrated that curacin A has an incredibly tight tubulin binding, with an immeasurable rate of dissociation. Though the K_d suggested irreversible binding, denaturation of tubulin enabled release of the small molecule, indicating that a covalent modification does not occur. Curacin A does not chemically resemble any other molecules in this subclass of tubulin inhibitors, superimposition of MM2 calculated energy-minimum configurations revealed structural similarities that represent a unique binding mode for colchicine-site occupying antimetotics (**33**).

The structure-function relationship of the Curacin A was well studied by Wipf and coworkers (**34, 35**), who synthesized more than 60 analogues and examined each for tubulin polymerization inhibition (TPI₅₀), competitive colchicine binding inhibition (CBI) and cell growth inhibition (CGI₅₀) in MCF-7 cells. Inversion of the stereochemistry of the cyclopropane, reduced the inhibition of colchicine binding by 46% as well as increased both TPI₅₀ and CGI₅₀. Similar loss of activity was observed when cyclopropyl moiety was replaced with a tert-butyl functionality, however substitution with 1-butylene showed slightly decreased CBI, but a 7.5-fold increase in CGI₅₀. Interestingly, inversion of the stereochemistry of the diene core, especially at the C-3 position *Z*-olefin, or chemical fragments of either the 2-methyl-cyclopropyl-thiazolinyl or the alkyl portions of the molecule, show an almost complete ablation of function (Table 1).

Table 1. Natural and synthetic analogs of curacin retain bioactivity.

Compound	ID	CBI, % at 5 μ M	TPI IC ₅₀ , μ M	GI ₅₀ MCF-7, 5 μ M
	1	94	0.72	0.04
	4	1	4.6	3.3
	5	83	0.92	0.30
	16	3	5.5	>1
	18	48	2.1	0.36
	19	0	>80	>10
	21	Inactive	Inactive	-
	25	35 \pm 2	100	0.13

1.4 Curacin biosynthesis The biosynthetic origins of this intriguing natural product were collaboratively elucidated by the Sherman and Gerwick groups, a decade following its initial reporting (**12**). Isotopic feeding studies, revealed that the molecule was derived from ten acetate units, two S-adenosyl-methionine derived methyl groups and one cysteine, suggesting that curacin A was the products of a mixed PKS/ NRPS system. Indeed, screening of a cosmid library prepared from genomic *Lyngbya majuscula* DNA by shotgun sequencing led to the localization of an operon containing 15 open reading frames, *curA-curN*. As a result of genome reassembly following 454 sequencing, the pathway was later revised (**36**), excluding the terminal module giving rise to the ~63kbp complete pathway consisting of ten PKS modules, and one NRPS module (Figure 6). Deviations of the structure of the mature metabolite that were predicted based on pathway translation prompted interest into biochemically characterizing the enzyme responsible for two unique elements at either termini of curacin A.

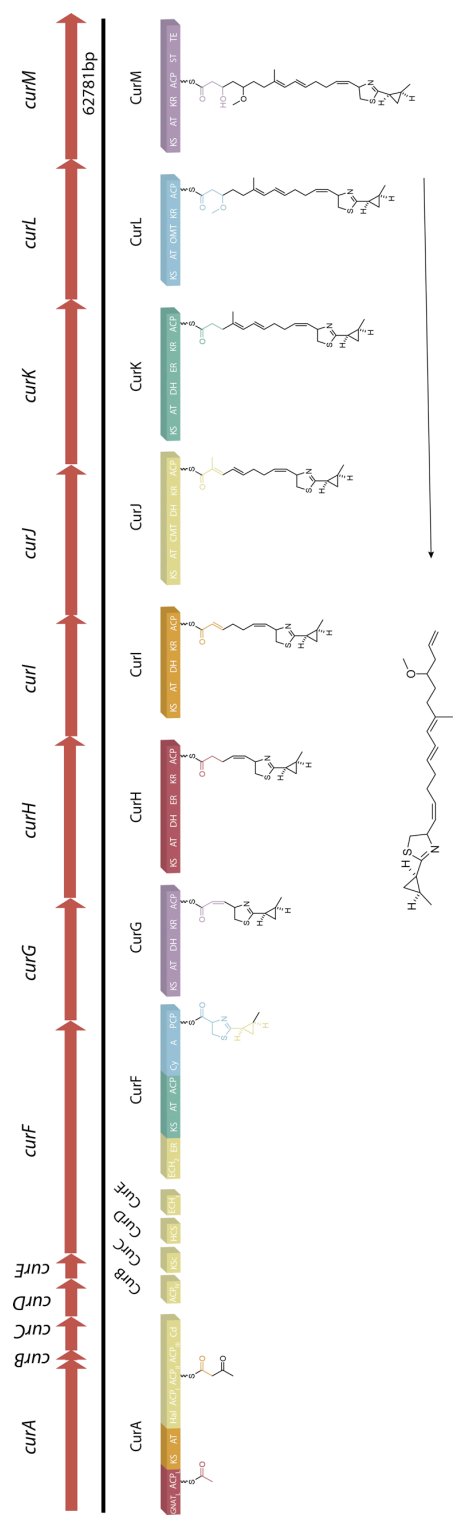


Figure 6. Curacin A biosynthetic pathway.

1.4.1 β -branching cassette leads to 2-methyl-cyclopropyl ring formation.

Inspection of the curacin pathway reveals four discrete domains between the first two PKS modules that share high sequence homology with proteins catalyzing the unusual isoprenoid-like β -branching transformations (37). This insertional biosynthetic array consists of 6 HCS cassette proteins hallmarks, 1) a stand alone ACP, 2) decarboxylating KS, 3) a 3-hydroxy-3-methylglutaryl-CoA synthase (HCS) and two individually encoded enoyl-CoA hydratases (ECH), 4) one catalyzing dehydration, 5) and a second catalyzing decarboxylation and 6) an ACP multiplet. Unique to the curacin and the very closely related jamaicamide pathways are the incorporation of a halogenase (Hal) domain, and an unusual C-terminal flanking region (Cd) downstream of the tandem ACP array (38).

Briefly, in the curacin system, CurD catalyzes the formation of HMG by condensing a CurB bound acetate, and the acetoacetyl-ACP_{I+II+III}, derived from the loading and PKS catalyzed elongation of CurA. The HMG-ACP_{I+II+III} reengages with CurA to undergo halogenation, yielding 4-chloro-HMG-ACP_{I+II+III}. This is followed by ECH₁, CurE, catalyzed dehydration giving rise to the 4-chloro-3-methylglutaconyl-ACP_{I+II+III} and subsequently CurF-ECH₂, catalyzed decarboxylation producing 4-chloro-3-methylcrotonyl-ACP_{I+II+III}. Cyclization is catalyzed by the unusual activity of the CurF-ER, which transfers hydride from NADPH, to the β position, leading to double bond electron rearrangement forming a new C-C bond and displacing the Cl⁻ (Figure 7) (39, 40).

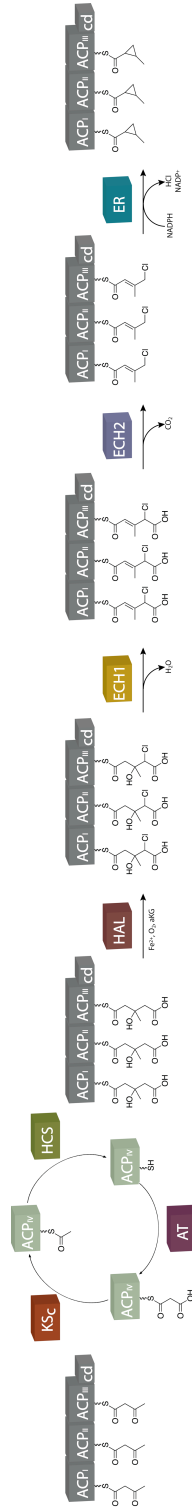


Figure 7. Curacin A β -branching assembly pathway.

An investigation into the role of the unusual nonheme-Fe²⁺ and α -ketoglutarate dependent halogenase, demonstrated that chlorination served as a crucial point of introducing structural diversity during curacin biosynthesis (40). It was found ECH₁-ECH₂ coupled catalyzed dehydration-decarboxylation of 4-Cl-HMG-ACP was 4-fold faster than the non-chlorinated substrate. Moreover, in the absence of the chlorine group, reaction with the ER catalyzed exclusive reduction of the 3-methylcrotonyl-ACP to the 3-methylbutanyl-ACP, thereby confirming chlorination to be indispensable for cyclization.

1.4.2 Activation-elimination offloading strategy leading to olefin formation

The biosynthetic origins of the unusual terminal alkene found in curacin A was illuminated by Gu et al. in 2009 (36). CurM includes a sulfotransferase (ST) domain integrated within the ACP and TE of the final module. STs utilize 3'-phosphoadenosine-5'-phosphosulfate to catalyze sulfonation of alcohols or amines. Within the context of the curacin pathway, the ST generates the β -sulfonic acid, from the corresponding KR produced (*R*)- β -alcohol. Subsequent desulfonative-decarboxylation, catalyzed by the unusual function of the TE leads to alkene formation (Figure 8). The use of a hydrocarbon-like substrate during this investigation suggested the capacity of these enzymes to be harnessed for liquid fuel production (36).

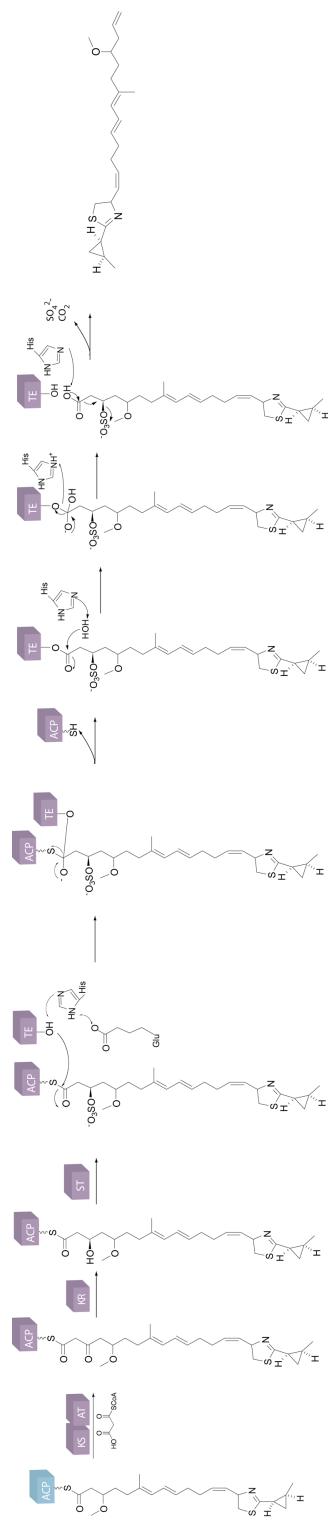


Figure 8. Curacin A olefinative offloading.

1.5 Research outline

Holistically, diversity in bioactive small molecules obtained from environmental sources, stems from unique modular architecture of a given pathway, which is emblematic of a type of naturally occurring “combinatorial biosynthesis”. In this mode, a library of related small molecules can be generated from a single gene cluster enabling trends in structure and activity to be determined and evolutionarily optimized. It stands then as a goal of natural product enzymologists, to make use of these biochemical processes in order to generate rationally designed, *de novo* small molecules *in vivo*. One can imagine a system in which a chimeric biosynthetic pathway is assembled and used to produce a protein templated route, giving rise to an engineered “unnatural product”. Ultimately what will enable us to achieve this goal, is an understanding of not only the standard biochemical “enzyme-ligand” domain-level interactions between substrate and catalytic partner but globally for enzyme megasynthases, an appreciation of the reigning regulatory mechanisms that control precise inter- and intramodular forces responsible for maintaining timing and order of protein communication throughout biosynthesis.

Herein, are described two studies of noncanonical megasynthase biocatalytic transformations observed during the construction of curacin A as well as an initial survey of the biosynthetic origins of the cyanobacterial metabolite carnabin A. Chapter two focuses on the unique β -branching protein-protein interactions leading to the pharmacophore assembly. The tandem carrier protein array as well as unusual C-terminal flanking domain are examined for their role in enhancing total biosynthetic turnover. The third chapter covers research concerned with expanding modular enzyme function toward the development of a novel biofuels platform. Guided by the initial investigation of the terminal curacin module, the ACP, ST and TE domains from CurM were examined for substrate promiscuity and catalytic activity. In addition, excised domains from a standalone cyanobacterial cluster containing shared ST-TE domain architecture were also studied. Chapter four includes a bioinformatic analysis of the newly identified pathway and partial biochemical characterization of cellular machinery responsible for alkyne formation and transaminative off-loading.

Chapter 2

Noncatalytic C-terminal protein domain enhances turnover during curacin β -branching

2.1 Introduction

Natural product structural diversification is typically achieved through the incorporation of variable elongation units and the downstream processing of the immature metabolite core (8). During polyketide synthase (PKS) biosynthesis, chemical elaborations are typically introduced into the final molecule through incorporation of functionalizations at the C-2 carbon of the incoming extender unit. Several natural product systems have evolved a unique set of off-pathway proteins that utilize the formation of a 3-hydroxy-3-methylglutaryl (HMG) group to broadly expand chemical structure (37) (Figure 9). This protein cassette is defined by a core set of six proteins including an acyl transferase (AT), acyl carrier protein (ACP_{IV}), keto synthase (KS) and HMG-CoA synthase (HCS), and two enoyl CoA hydratases (ECH). Typically a carrier protein multiplet, consisting of two or more domains, is also observed.

During elongation, the stand-alone AT transfers a malonate from malonyl-CoA to the corresponding ACP, followed by KS catalyzed decarboxylation. The HCS catalyzes bifunctional condensation of an acetate group at the β -position substrate bound to the tandem ACP, followed by hydrolysis yielding the HMG. The site specificity of the formation of this functional group provides the nomenclature for this type of “ β -branching”. The HMG undergoes successive dehydration and decarboxylation catalyzed by individual ECH domains. This moiety can be further processed and incorporated into the structure of the mature compound.

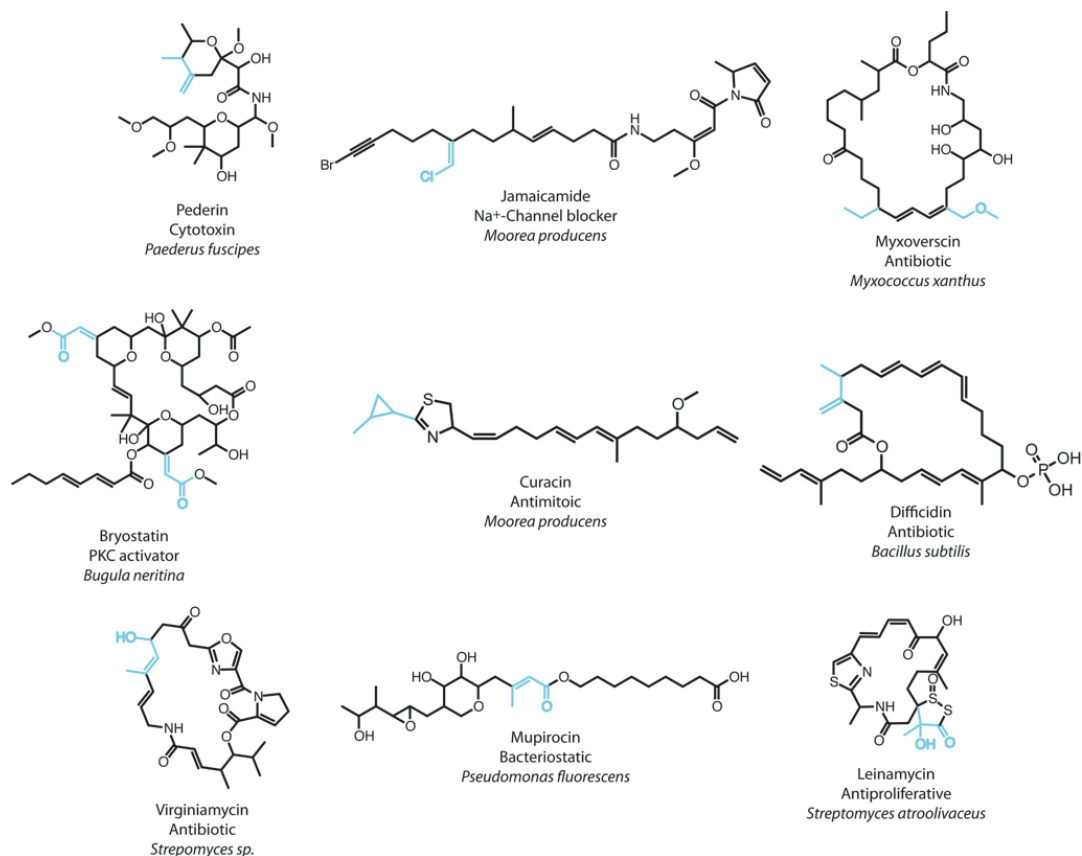


Figure 9. Natural products harboring β -branching functionalization. Blue highlight indicates structural element introduced by the protein cassette.

Previously, the formation of the 2-methylcyclopropyl group found in curacin A (Cur) and the pendant vinyl chloride from the related jamaicamide (Jam) pathways were found to originate from an embedded HCS cassette in these two highly homologous pathways from the marine cyanobacterium *Moorea producens* (40). Interestingly, in addition to the HCS cassette, both systems, contain an embedded halogenase, which in the curacin pathway has been shown to be indispensable for the formation of the cyclopropane (Figure 10). Also unique to these pathways is a large 8kDa-flanking domain C-terminal (Cd) of the ACP triplet. The Cd was demonstrated to increase total yield of a reconstituted β -branching assembly, by a greater than 2-fold increase. Experiments parsing the mechanism of this enhanced biosynthesis are the focus of this chapter.

2.2 Results

The highly conserved, HMG-CoA Synthase (HCS) assembly found in the curacin pathway was previously shown to produce the pharmacophoric 2-methyl-cyclopropyl moiety (**38**). Briefly 3-hydroxy-3-methyl-glutaryl-ACP is formed by a stand-alone HMG-CoA synthase by condensing acetyl-ACP and acetoacetyl-ACP. Chlorination at the γ -position catalyzed by the embedded Halogenase (Hal), domain gives rise to the chlorinated substrate, which is critical for further modification by a dehydrating enoyl-CoA hydratase (ECH₁), decarboxylating enoyl-CoA hydratase (ECH₂). Finally the modified enoylreductase (ER) catalyzes cyclopropane formation (Figure 10). Together these proteins constitute the isoprenoid-like β -branching assembly.

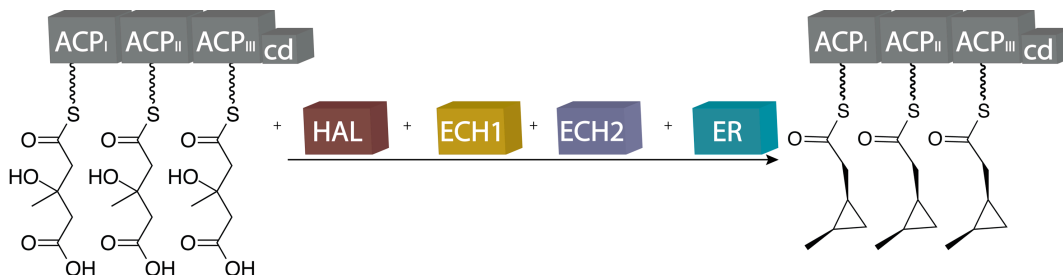


Figure 10. Curacin A 2-methyl-cyclopropane is formed by β -branching protein. Cyclopropanation is the result of interactions between the ACP_{I-II-III-Cd}, and Hal, ECH₁, ECH₂, and ER. See text for domain abbreviations.

A hallmark of this type of insertional cassette, also found in the bryostatin (**41**), viriginamycin (**42**), pederin (**43**) and bacillaene systems (**44**), is the presence of carrier protein multiplet, contrasted with canonical PKS and NRPS architecture, in which only a single thiolation domain, ACP or PCP, is present per module. Unique to the β -branching assemblies in the Cur and Jam systems, which possess a carrier protein triplet with 94% sequence identity (Figure 11), ACP_{I+II+III}, is the presence of 8 kDa C-terminal flanking domain (Cd), 71.2% pairwise identity (Figure 12), giving rise to the unusual tetradomain ACP_{I+II+III+Cd}.

Consensus	1	10	20	30	*	40	50	60	70	74	
1. JamE ACP3Cd - ACPI	V	KQVLKKEQLA	EALYTEES	SEIAEDQKFV	DLGLDS	IVGV	EWTTTINQ	TYNLN	LKATKLY	DYPTLLE	LAEYTAQTL
2. JamE ACP3Cd - ACPII	E	IKQVLKKEQLA	EALYTEES	SEIAEDQKFV	DLGLDS	IVGV	EWTTTINQ	TYNLN	LKATKLY	DYPTLLE	LAEYTAQTL
3. CurA ACP3Cd - ACPII	E	IKQVLKQQLA	EALYTEES	SEIAEDQKFV	DLGLDS	IVGV	EWTTTINQ	TYNLN	LKATKLY	DYPTLLE	LAAEYTAQTL
4. CurA ACP3Cd - ACPII	I	KQVLKQQLA	EALYTEES	SEIAEDQKFV	DLGLDS	IVGV	EWTTTINQ	TYNLN	LKATKLY	DYPTLLE	LSGYTAQTL
5. CurA ACP3Cd - ACPIII	E	IKQVLKQQLA	EALYTEES	SEIAEDQKFV	DLGLDS	IVGV	EWTTTINQ	TYNLN	LKATKLY	DYPTLLE	LAPYTAQEI
6. JamE ACP3Cd - ACPIII	E	VKQVLKQQLA	EALYTDI	SEIAEDQKFV	DLGLDS	IVGV	EWTTTINQ	TYNLN	LKATKLY	DYPTLLE	LSAYTAQEI

Figure 11. Sequence alignment of excised ACP domains from the CurA and JamE carrier protein multiplets. Asterisk indicates the invariant, post-translationally phosphopantetheinylated serine.

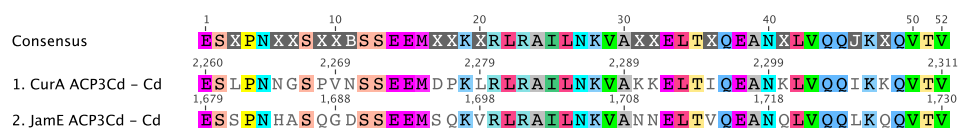


Figure 12. Sequence alignment of the terminal C-terminal flanking domains, Cd, from the CurA and JamE carrier protein multiplets.

2.2.1 Cd increases biosynthetic flux of a reconstituted β -branching system

In order to determine the functional role of the Cd domain, a series of carrier protein constructs were prepared. Proteins were designed varying the number of ACPs as mono-, di-, tridomains, as well as Cd truncations lacking the terminal domain. Additionally, carrier protein knockouts, with the active site serine mutated to alanine preventing substrate loading were also generated, in order to examine whether the ACPs are capable of functioning in parallel or whether transformation occur in series, as observed in polyunsaturated fatty acid biosynthesis (45). Proteins were overexpressed, purified and loaded with (*S*)-HMG to generate the (*S*)-HMG-ACP substrate. The carrier proteins were incubated in one-pot reaction along with the Hal, ECH₁, ECH₂, ER and NRPS domains from CurF. As a metric of complete processing of the HMG by both the β -branching enzymes as well as NRPS module, 2-methyl-cyclopropyl-thiazolinic acid was released from the PCP via base catalyzed hydrolysis and quantified via LC-MS (Figure 13) (38).

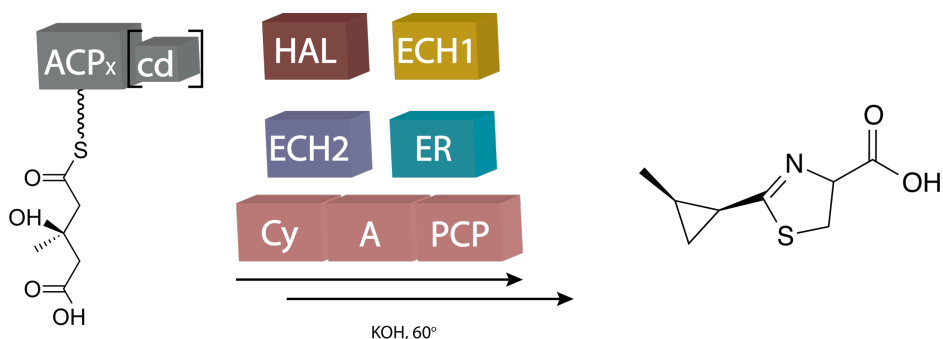


Figure 13. Schematic for one-pot reaction and metabolite analysis. Characterization of β -branching interaction and effect of ACP isoform on 2-methyl-cyclopropyl-thiazolinic acid formation.

For a single excised carrier protein, approximately 10% conversion was observed. Doubling, the number of ACPs, prepared as either ACP_{I+II} or ACP_{II+III}, yielded a stoichiometric increase in conversion, 18-24%, while the tridomain, ACP_{I+II+III}, led to 37% overall yield. For the ACP

knockouts, yields of thiazolinic acid were proportional to the number of competent ACPs; double mutants, converted the substrate to near 10%, while the single mutants performed between 19-25%. These numbers were consistent with excised mono- and didomain constructs indicating that the ACPs were capable of functioning independently. Unexpectedly, the addition of the Cd domain to the ACP triplet, ACP_{I+II+III+Cd}, was capable of increasing the conversion of HMG-ACP to 2-methyl-cylcopropyl-thiazolinic acid 2.7-fold to 100%, relative to the construct lacking the Cd. Similarly, the rate enhancement was also observed for ACP double knockouts possessing the Cd with increased rates of conversion 2.6-3.0-fold when the Cd was present.

2.2.2 Biophysical characterization of Cd on protein dynamics

Our first insight into how these proteins might be acting in order to increase the total turnover came during the purification process. ACP_{I+II+III} and ACP_{I+II+III+Cd} were prepared as highly concentrated samples and were loaded onto a Superdex Gel filtration column in order to further purify protein samples by size exclusion chromatography (SEC). Unusually, both proteins eluted at a volume corresponding with a size of approximately 1.25-2 the expected molecular weight (Figure 14). Analytical gel filtration confirmed that ACP_{I+II+III} had an apparent molecular weight of 63.1 ± 5 kDa, while the ACP_{I+II+III+Cd} had eluted at a volume corresponding with 106.6 ± 7 kDa.

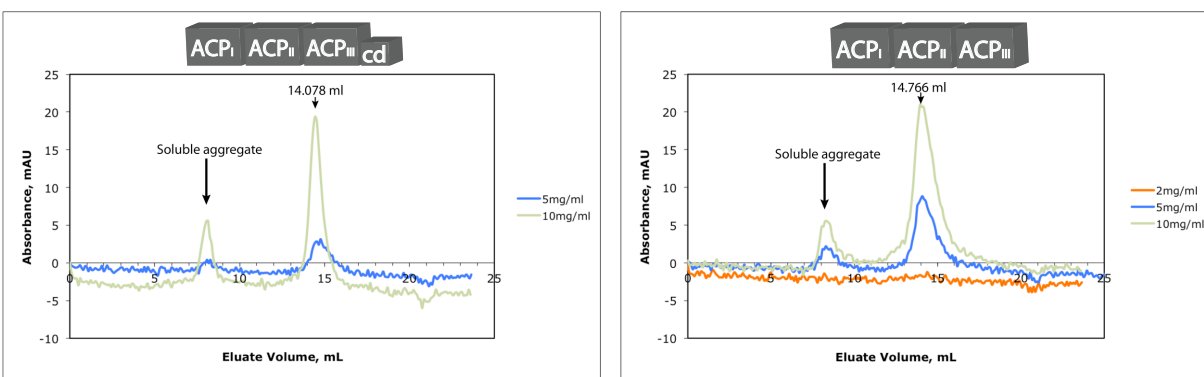


Figure 14. Analytical size exclusion chromatography for CurA ACP multipliants. Concentration independence of apparent molecular weights increases for ACP_{I+II+III} and ACP_{I+II+III+Cd}. Traces were recorded at 280nm. Green trace- 10mg/ml; Blue trace- 5mg/ml; Red trace- 2 mg/ml.

Excised ACP constructs showed similar increases in apparent Concentration molecular weight (Table 2). Though differences between observed and calculated molecular weights increased with the number of ACP domains present, the greatest proportional increase was affected by the presence of the 8kDa Cd. This phenomenon appeared to be concentration independent, as no

change in the apparent molecular weight for either $ACP_{I+II+III}$ or $ACP_{I+II+III+Cd}$ was observed over a range of sample concentrations (Figure 14).

Table 2. Apparent molecular weights of ACPs by analytical gel filtration chromatography.

ACP construct	Calculated MW, kDa	Observed MW, kDa
ACPI	10.9	13.1± 0.5
ACPII	11.3	14.2± 0.5
ACPIII	11.0	13.2± 0.7
ACPI-ACPII	23.6	42.3± 2.5
ACPII-ACPIII	22.6	40.7± 2.0
ACPI-ACPII-ACPIII	36.0	63.1± 5.1
ACPI-ACPII-ACPIII-Cd	44.6	106.6± 7.3

These findings suggested that unlike ACPs from canonical PKS or fatty acid synthase systems, which are monomeric (46), the unique tandem carrier proteins found in the Cur β -branching cassette may exist in equilibrium with a higher order protein oligomer.

The apparent stabilization of a Cd mediated, ACP-triplet quaternary structure, was further investigated using, biophysical methods. Analytical ultracentrifugation (AUC) has been used to determine the shape and binding stoichiometry of proteins by monitoring sedimentation of the macromolecule under high centrifugal force, over time (47). The purified $ACP_{I+II+III}$ and $ACP_{I+II+III+Cd}$ were prepared at three different concentrations and analyzed independently by sedimentation velocity. By AUC both ACP constructs appeared to be prolate ellipsoids, with frictional coefficients ranging from 1.47-1.53 (Table 3). The observed molecular weights were determined to be 36.7 ± 0.8 kDa and the 42.0 ± 1.9 kDa, for $ACP_{I+II+III}$ and $ACP_{I+II+III+Cd}$, respectively, indicating that both proteins were monomeric in solution.

Table 3. Biophysical constants obtained for CurA carrier proteins by analytical ultracentrifugation.

Protein Construct	ACP _{I-II-III}	ACP _{I-II-III-Cd}
Sed. Coeffi., 10 ⁻¹³ s	2.58± 0.04	2.74± 0.03
f/f(0)	1.47± 0.02	1.53± 0.05
Observed MW, Da	36788± 830	42035± 1964
Calculated MW, Da	36702	44614

Though this finding is incongruous with the results of the SEC, previous investigations into protein stoichiometry have demonstrated the susceptibility of weak oligomeric interactions to be disrupted by the high hydrodynamic pressure generated by AUC conditions (48); thus we sought to undertake a direct inspection of the oligomeric state of the protein by electron microscopy (EM). ACP constructs were purified to homogeneity by preparative gel filtration and diluted to a uniform concentration. Samples were then loaded on a carbon grid and overlaid with uranyl formate to enable visualization by negative staining. Individual particles were selected, binned and sorted according to class averages (Figure 15). For the ACP_{I+II+III}, the protein population was determined to be 74% monomeric and 25% dimeric, while for the ACP_{I+II+III+Cd}, the distribution was 29% monomeric and 70% dimeric. This mixed population of ACPs detected by EM supported the hypothesis that the Cd mediated shift in the ACP self-association equilibrium towards the dimer, was a possible explanation for the observed increase in yield.

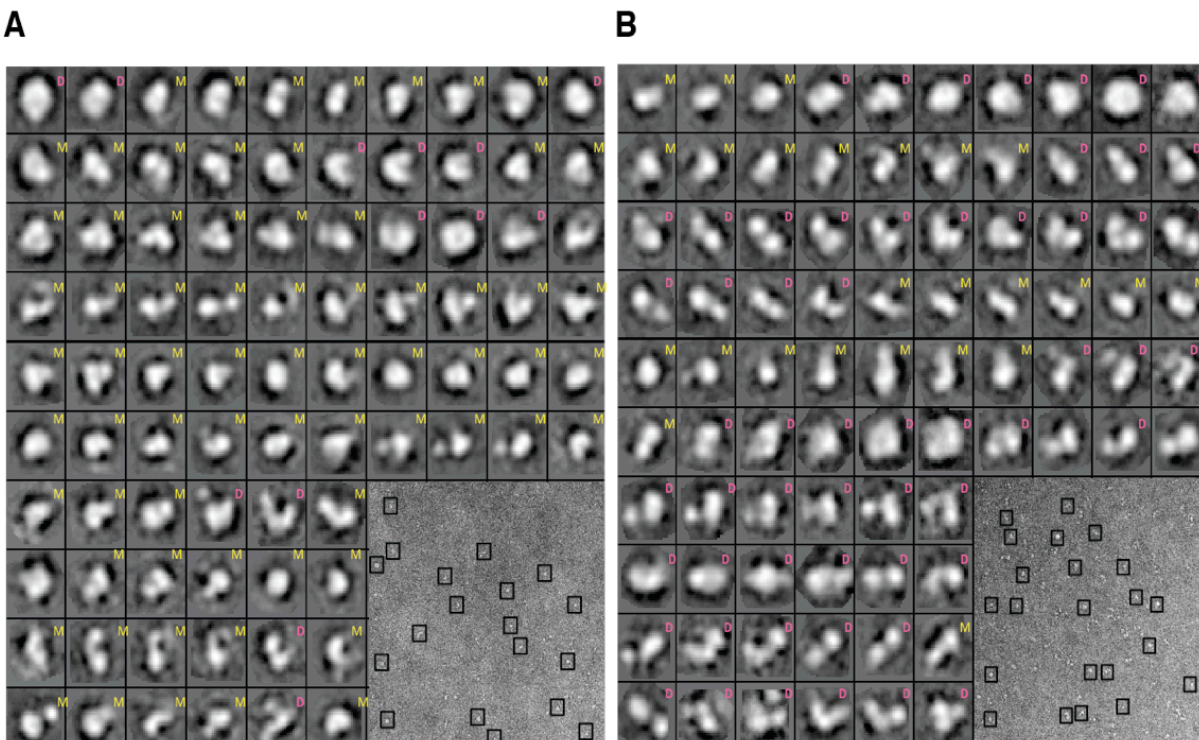


Figure 15. Electron microscopy analysis of CurA ACP multiplsets by negative staining. 2D class averages and raw data (inset) for A. ACP_{I-II-III} and B. ACP_{I-II-III}-Cd.

2.2.3 Preliminary characterization of Cd catalyzed self-association

As demonstrated by negative staining, the Cd shows an effect on shifting the protein equilibrium of the ACP triplet towards dimerization. In order to calculate the binding constants afforded exclusively by the Cd, the 8kDa region was cloned and overexpressed as an excised domain. This purified fragment was concentrated to >600 μM , in order to obtain as much of the Cd-dimer as possible. Dilution experiments were carried out by isothermal titration calorimetry (ITC) with the Cd, by injecting small volumes of the concentrated protein into a sample-cell containing only buffer. Heats of dilution were plotted against mole fraction buffer and the curve was fitted to a simple two state equilibrium with an apparent K_d of 1.96 ± 0.11 mM (Figure 16).

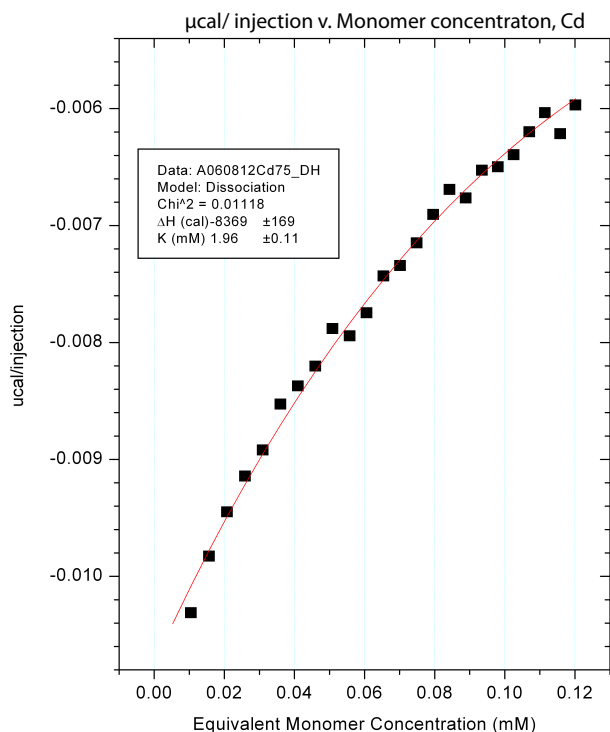


Figure 16. ITC dilution assay for the CurA Cd. Heats of injection of concentrated Cd dimer, diluted into buffer, and plotted against fraction monomer.

The extraordinarily high value of this dissociation constant, which extends beyond the experimental range, suggested that additional domains might be necessary to observe Cd functioning as a dimerization domain. To examine this, wild type constructs of the $ACP_{I+II+III}$ and the $ACP_{I+II+III+Cd}$ were prepared and concentrated to $>200 \mu\text{M}$, a concentration above which the protein is prone to precipitation. Again, dilution experiments were performed with K_d values of $1.78 \pm 0.2 \text{ mM}$ and $1.88 \pm 0.2 \text{ mM}$ obtained. While the absence of significant heats of dilution were consistent with the previously obtained results for the excised Cd domain, the origin of the enhanced binding affinity observed during the EM analysis remained elusive.

2.2.4 Mechanistic insights into cyclopropane formation by individual reaction monitoring

Through negative staining and analysis by EM, we demonstrated that the Cd was capable of shifting the oligomerization equilibrium of the CurA ACP triplet towards the dimer. However, we sought to examine the physiological role of the Cd in facilitating enhanced biosynthetic flux

by looking at discrete reactions along the Cur β -branching pathway.

Interactions between megasynthases are dictated by short, protruding regions adjacent to the C-terminus and N-terminus of the donor and acceptor modules, respectively. PKS docking domains (dd) and NRPS communication domains (com) maintain not only specificity of the modular protein-protein interactions but also efficient transfer of substrate (49-52). Global alignment of the docking domains within the Cur pathway revealed two unique docking interfaces between the β -branching modules, CurA-CurF, as well as between CurF-CurG at the NRPS-PKS interface. Such conspicuous dissimilarities in sequences, led to the hypothesis that the Cd might function in a novel, β -branching specific, docking mode. While the Cd has only been identified in Cur and Jam systems, sequence alignment of β -branching modules that terminate in an ACP multiplet revealed a 34 amino acid region with 27.8% sequence similarity for the C-terminus of the donor module, and 24 N-terminal amino acids with 36.4% sequence similarity derived from the ECH₂ of the down stream acceptor module, supporting the role of the Cd in modular docking interactions.

To examine the role of the Cd in enabling specific protein interactions, constructs of the terminal carrier protein lacking or possessing the Cd, ACP_{III} or ACP_{III+Cd}, respectively, were generated. ACPs were overexpressed and purified in the apo-form and confirmed by MS. Proteins were then loaded with (*R, S*)-HMG-CoA using the phosphopantetheinyl transferase, Sfp. Individual reactions were monitored stepwise by protein mass spectrometry. Herein, reaction aliquots were quenched at various time-points with the addition of the formic acid at a concentration sufficient to denature, but not precipitate the proteins. Unlike the previous investigation of the β -branching assembly however, the substrate was not hydrolyzed from the ACP, rather the proteins were subject to inline LC-MS in order to analyze modifications to the HMG substrate (Figure 17). Additionally, as the preferred stereochemistry of the Hal and ECH₁ substrates are known to be (*S*)-HMG, maximal rates of conversion are expected to be 50% resulting from racemic mixture used to load the carrier proteins.

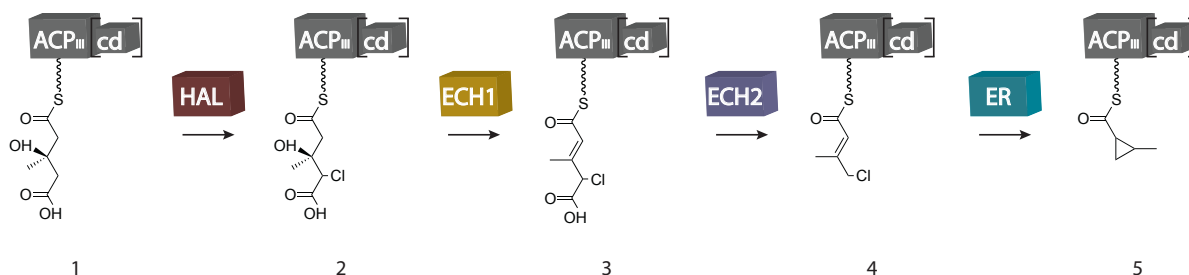


Figure 17. Examination of discrete β -branching interactions by LC-MS. Schematic of individual or combinations of reactions catalyzed upon incubation of HMG-ACP with excised domains. Rates of conversion were determined by both deconvoluted mass abundance as well as comparison of charge states for ACP modifications

2.2.4.1 Cd dramatically increases rate of halogenation

We have previously demonstrated that Hal requires Fe^{2+} and $\alpha\text{-KG}$ as cofactors to generate γ -chloro-HMG. In order to provide holo-halogenase, Hal was incubated with freshly prepared $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ and $\alpha\text{-KG}$ prior to the addition of ACP-loaded substrate. The reactions were initiated with the addition of HMG-ACP, and extent of chlorination was analyzed by mass spectrometry. For the excised ACP_{III} construct, chlorination occurs, with 26.0% conversion by 5 minutes, and only 45.1% converted in 30 minutes. Remarkably, for the $\text{ACP}_{\text{III}+\text{Cd}}$, 27% is chlorinated within 30 seconds and the reaction is complete after 3 minutes (Figure 18).

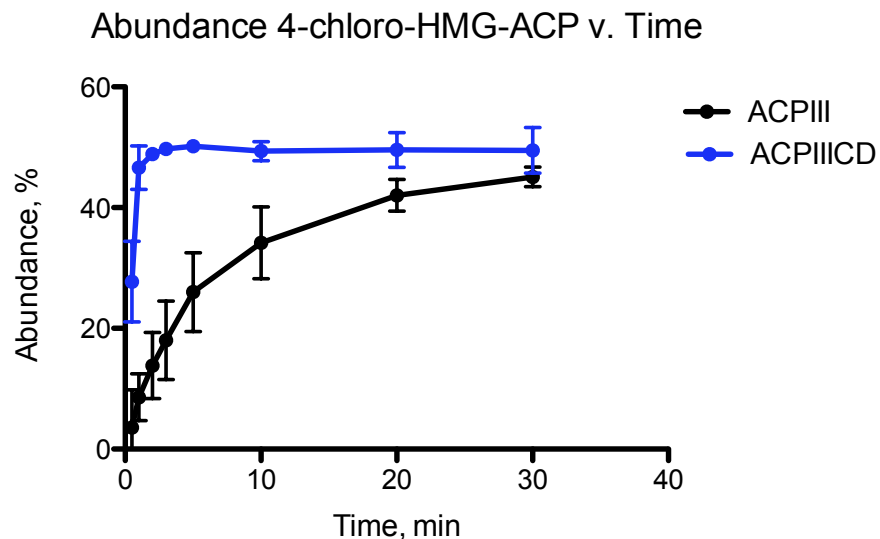


Figure 18. Cd dramatically increases rate of halogenation of HMG-ACP. Black circles- ACP_{III} . Blue circles- $\text{ACP}_{\text{III}+\text{Cd}}$

2.2.4.2 Cd does not affect rate, but increases extent of dehydration

To obtain 100% concentrations of the 4-chloro-HMG-ACP substrate for dehydration, the reactions were prepared as for the halogenation reaction, and allowed to react for 45 minutes prior to the addition of the ECH₁. Difficulties in obtaining sufficiently pure and active CurE ECH₁ led us to use JamI ECH₁ as a surrogate dehydratase. The rate of formation of 4-chloro-3-methylglutaconyl-ACP for both ACP isoforms was incredibly rapid and the maximal conversion achieved within 60 seconds FIG, though a slight increase in the total abundance of the dehydrated product is observed for the Cd containing construct, 14.0% versus 15.8% (Figure 19).

Abundance 4-chloro-3-methylglutaconyl-ACP v. Time

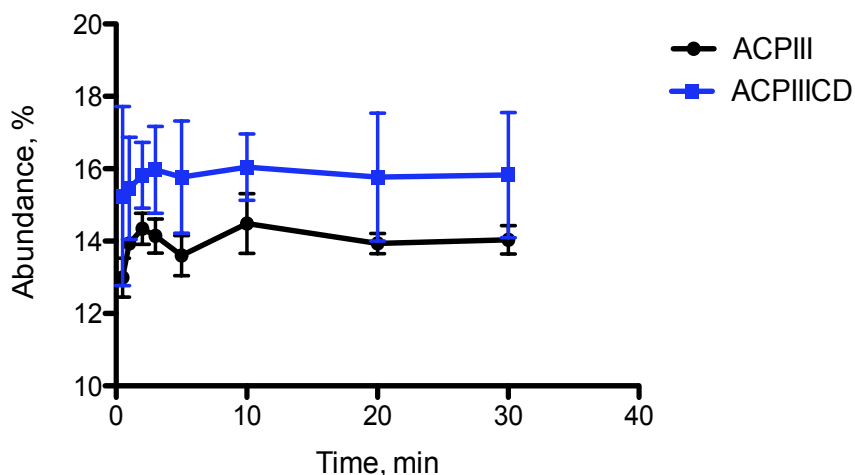


Figure 19. Effect of Cd on ECH₁ activity. Cd does not effect rate, but increases extent of dehydration of Black circles- ACP_{III}. Blue squares -ACP_{III}+Cd.

Unexpectedly, the majority of the 4-chloro-HMG-ACP remains unreacted for both the ACP_{III} and the ACP_{III}+Cd (Figure 20). The origins of apparent reaction stalling are unknown, though it is possible that the reaction is reversible and sensitive to the presence of the downstream catalytic proteins.

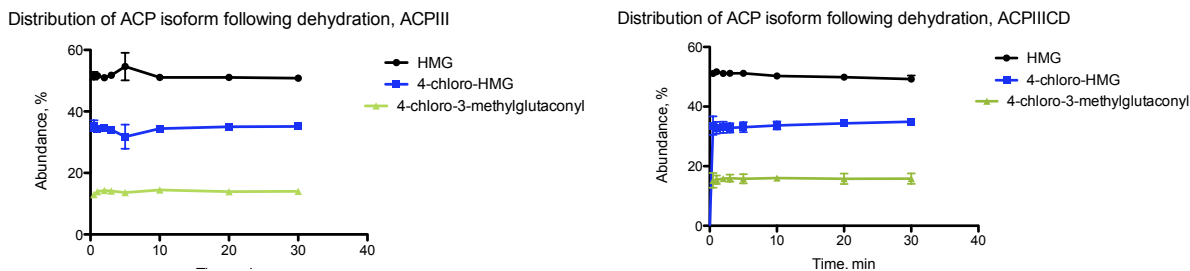


Figure 20. Evaluation of ACP isoform following ECH1 interaction. As evidenced by the presence of chlorinated substrate, dehydration is incomplete and independent of Cd. Black circles- HMG-ACP. Blue squares 4-chloro-HMG-ACP. Green triangles- 4-chloro-3-methylglutaconyl-ACP

2.2.4.3 Cd does not effect rate, but increases extent of decarboxylation

As with dehydration, no substantial differences in the rate of decarboxylation were observed (Figure 21) upon addition of the ECH₂WT, however 4-chloro-3-methylcrotonyl-ACP_{III+Cd} was 1.5-fold more abundant, 23.8% v. 36.0%, than for the ACP_{III} by the end of the reaction. Similarly with the ECH₁ catalysis, decarboxylation of substrate does not go to completion for either ACP construct

Abundance 4-chloro-3-methylcrotonyl-ACP v. Time

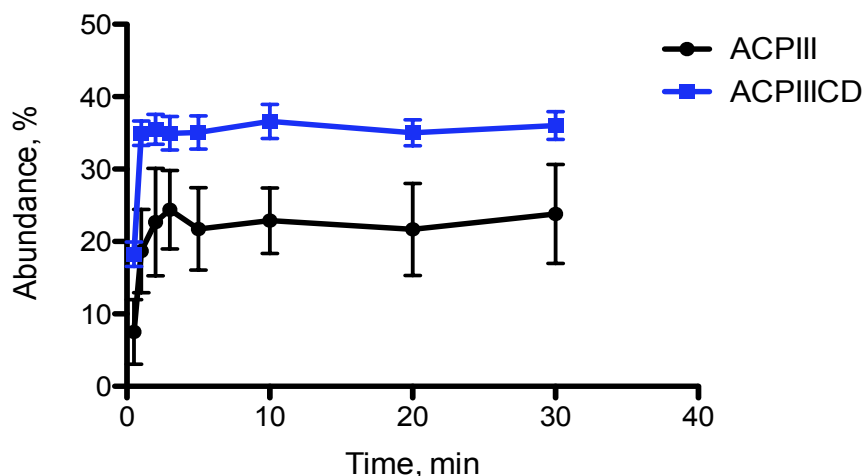


Figure 21. Effect of Cd on ECH2 activity. Extent of decarboxylation is incomplete and dependent of Cd. Black circles- ACP_{III}. Blue squares -ACP_{III}-Cd.

2.2.4.4 ECH2 N-terminus has minimal effect on rate of decarboxylation

To determine the potential effect of the CurA-CurF docking interaction, an N-terminally truncated ECH₂dd was generated lacking the first 15 amino acids. For the ACP_{III}, a slight increase in the total amount of the 4-chloro-3-methylcrotonyl product was observed, however

this number is within error of that observed for the ECH₂WT, while no differences in either rate or extent of decarboxylation were observed for the ACP_{III+Cd} construct (Figure 22).

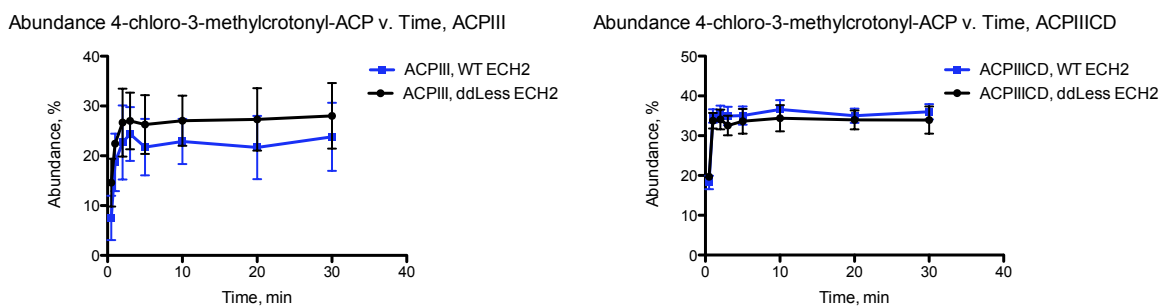


Figure 22. ECH₂ activity is independent of N-terminal docking domain and presence of Cd. Blue squares- WT ECH₂. Black squares- docking domain-less ECH₂

2.2.4.5 Cd does not effect rate, but increases extent of cyclization

The effect of the Cd on ER catalyzed cyclopropanation, followed a similar trend as observed with the remaining β -branching proteins. The maximal conversion of 2-methyl-cyclopropane was achieved rapidly for both the ACP_{III} as for the ACP_{III+Cd}, while the total yield was 1.3-fold higher for the Cd containing construct, 36.1% v. 47.0% (Figure 23).

Abundance 2-methylcyclopropyl-1-carboxyl-ACP v. Time

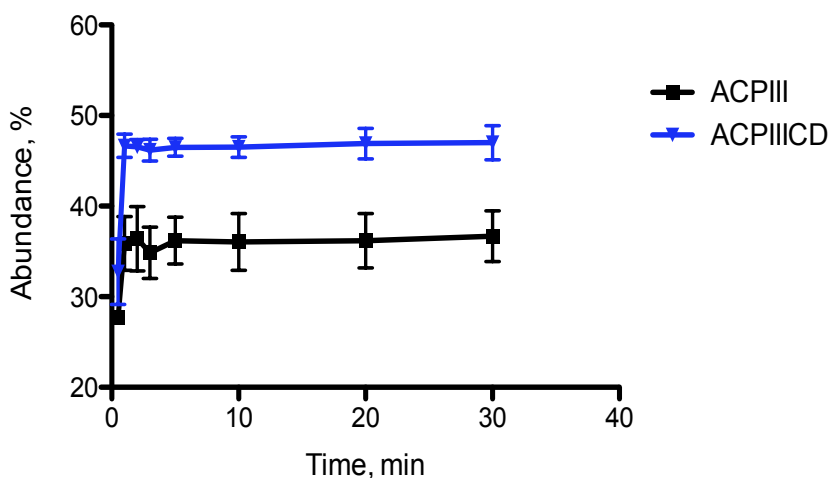


Figure 23. Effect of Cd on ER activity. Extent of cyclopropanation is incomplete and dependent of Cd. Black squares- ACP_{III}. Blue triangles -ACP_{III+Cd}.

2.2.4.6 Reconstitution of β -branching cassette in a one-pot reaction

In assessing the sum function of the enzymes in the conversion of HMG to 2-methylcyclopropane, the β -branching cassette was reconstituted in a one-pot reaction mixture. Hal, ECH₁, ECH₂WT, and ER were incubated for 10 minutes along with the necessary Hal cofactors as well as NADPH, required for ER functionality. Reactions were initiated with the addition of HMG-loaded ACP. Over the time course of the reaction, 1.2-fold greater cyclopropyl-ACP was formed for the Cd containing construct $35.3 \pm 1.3\%$, whereas $28.8 \pm 6.1\%$ was generated for ACP_{III} (Figure 24). Additionally, the rate of conversion is also faster of the ACP_{III+Cd}, suggesting halogenation is a critical rate-determining step. This trend of increased conversion is in agreement with what was previously observed for individually monitored reactions.

Abundance 2-methylcyclopropyl-1-carboxyl-ACP v. Time, one-pot assay

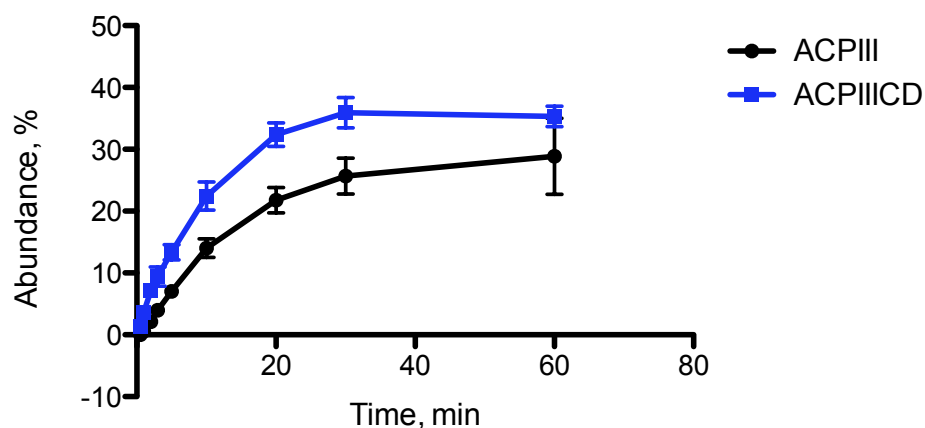


Figure 24. Effect of Cd upon β -branching reconstitution. Extent of cyclopropanation is evidenced by the increase in total abundance and rate of cyclopropane formation. Black circles- ACP_{III}. Blue squares -ACP_{III+Cd}.

Global analysis of all ACP isoforms (Figures 25, 26) over time revealed a rapidly formed substrate pool of the chlorinated and dehydrated species that is converted to the cyclopropane by the ER. For the ACP_{III}, $5.5 \pm 1.1\%$ $6.7 \pm 0.5\%$, of the chlorinated and dehydrated products were detected, while for the ACP_{III+Cd}, slightly more 4-chloro-HMG-ACP $9.5 \pm 0.5\%$, was detected compared with the Cd-less construct, while a moderate decrease in the dehydrated 4-chloro-3-methylglutaryl-ACP, $6.0 \pm 0.4\%$, was observable. Interestingly no decarboxylated product is

detectable for either ACP_{III} or ACP_{III+Cd} , suggesting the rapid conversion of the 4-chloro-3-methylcrotonyl species to the cyclized form.

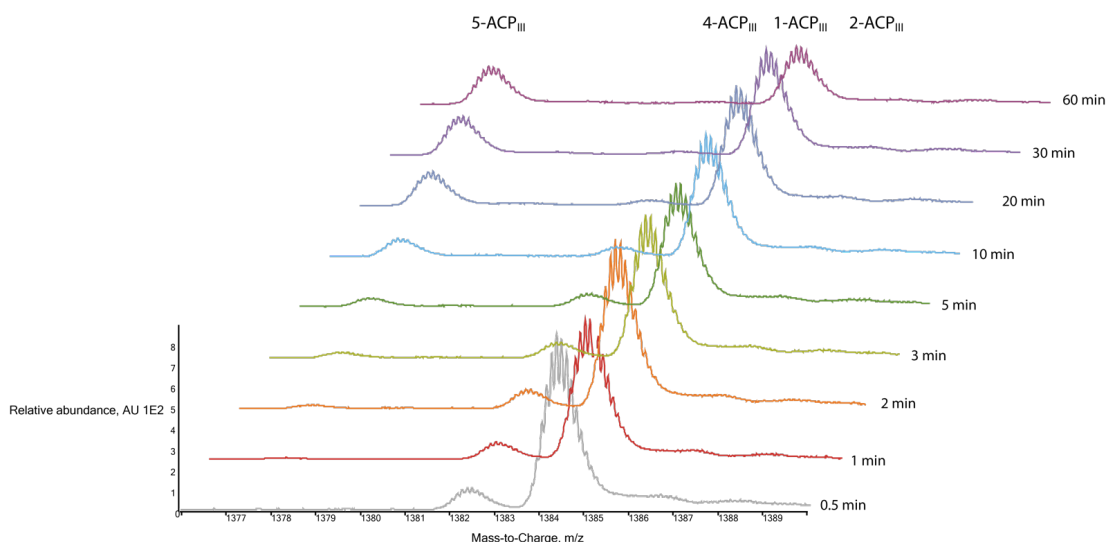


Figure 25. Distribution ACP isoform over time in the one-pot assay cyclopropanation assay, ACP_{III} , +9 charge state. See text for discussion.

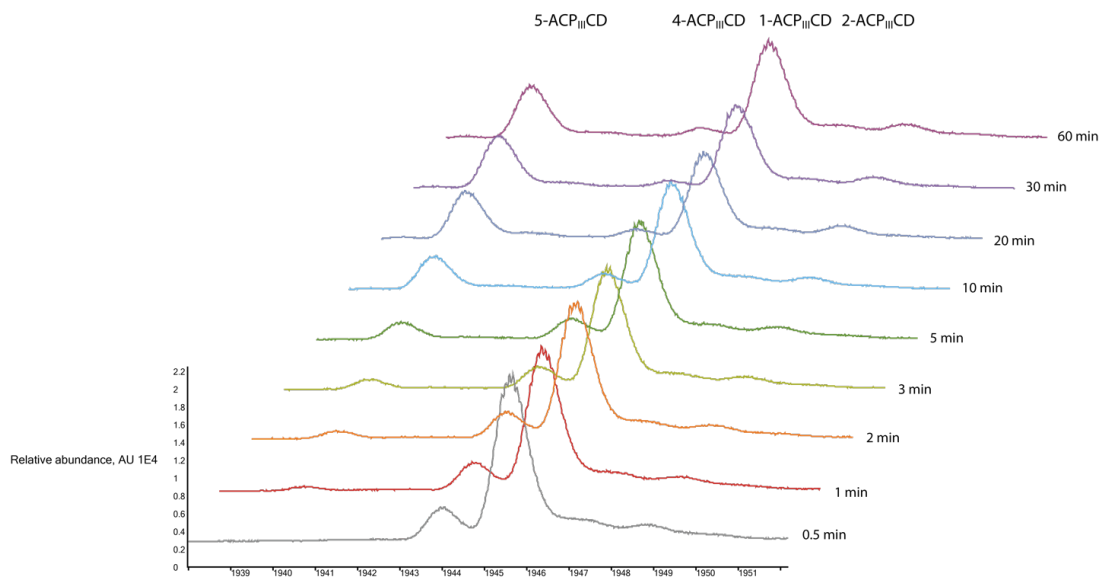


Figure 26. Distribution ACP isoform over time in the one-pot assay cyclopropanation assay, ACP_{III+Cd} , +11 charge state. See text for discussion.

From these experiments, a more complete understanding for cyclopropane formation can be developed, wherein individual domains each play a critical role in the biosynthesis of the unusual structural moiety (Figure 27).

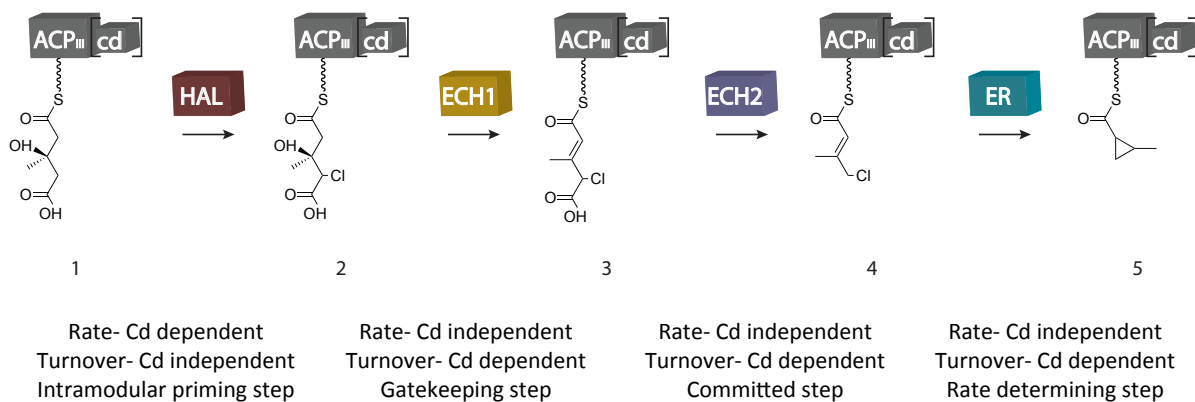


Figure 27. Role assignment for individual β -branching reactions during cyclopropane formation based on LC-MS analysis.

2.2.5 Solution structure of ACP_I and characterization of Hal-ACP interaction by MS

The unusually high sequence identity shared between the CurA and JamE ACP triplet, and the significant role of the unusual halogenase as part of the β -branching cassette in pharmacophore biosynthesis led us to structurally probe ACP dimerization and ACP-Hal interactions (53).

¹⁵N-labeled ACP_{I+II+III} and ACP_{I+II+III+Cd} were prepared and used to study the self-association of the carrier protein triplet. To account for the large size of the protein transverse relaxation optimized spectroscopy (TROSY), rather than heteronuclear single quantum coherence spectroscopy (HSQC) was employed, and [¹⁵N, ¹H]-TROSY spectra were recorded for both isoforms of the ACP triplet. Addition of Cd led to minor decreases in peak intensities overall, however no differences in chemical shifts for the core helices were detected, as compared with the construct lacking the domain. This result suggested that the ACPs adopt a uniform conformation independent of the presence of the Cd. In order to detect discrete domain-domain interactions resulting from protein dimerization, titration experiments were carried out using ¹⁵N-labeled excised ACP_I monodomain, or ACP_{I+II} or ACP_{II+III} didomain constructs, and incubated with ¹⁵N-ACP_{III} and similarly examined as for the full tridomain. To our surprise no perturbations in chemical shifts were observed for titration of any combination of the proteins, suggesting the intra-ACP-domain interactions do not occur.

The solution structure of the excised ACP₁ was solved by NMR and shown to be a right-handed twisted bundle, comprised of four α -helical domains. Typical, carrier proteins contain three helical domain, however the somewhat longer, highly hydrophobic interloop region between helices α II and α IV, was ordered and adopted a helical conformation. The unexpected lack of structural differenced of the ACPs when dimerized, prompted us to examine the potential for substrate mediated conformational changes that take place for an excised ACP. Indeed, conversion of the ACP from the apo- to the holo- form via *in vitro* phosphopantetheinylation showed the peak of the invariant active serine, Ser1989, has a substantial downfield shift from 9.0 to 8.5ppm suggesting modest structural rearrangement to accommodate the PPant prosthetic group. However carrier proteins subsequently prepared as, HMG-loaded or, 4-Cl-HMG-loaded, or as the nonnative acetyl-loaded, only exhibited minor shifts surrounding the PPant. Perturbations in chemical shifts were greatest for residues in proximity to the active site, specifically, A2009, V1983, I1990, D2014, and D1988, presumably to account for the increase in substrate bulk of the loaded as compared with the holo- protein.

An examination of these residues in enabling specific Hal-ACP interactions was carried out by mutagenesis and comparing the ratio of 4-chloro-HMG-ACP₁: HMG-ACP₁ for each mutant relative to wild type. Initially these measurements were collected by MALDI-TOF, however, signal broadening and peak overlap led us to utilize LC-ESI-QTOF (Figure 28). This technique enabled quantification of protein activity by simultaneous integration of HPLC peaks of the separated proteins, as well as by charge abundance for the chlorinated and non-chlorinated proteins.

Though differences in activity vary between analytical methods, consistently, the most deleterious point mutants were constrained to the area surrounding the active site serine. The most severe, D1998A, I1990A and A2009R, showed a greater than 70% decrease in chlorination.

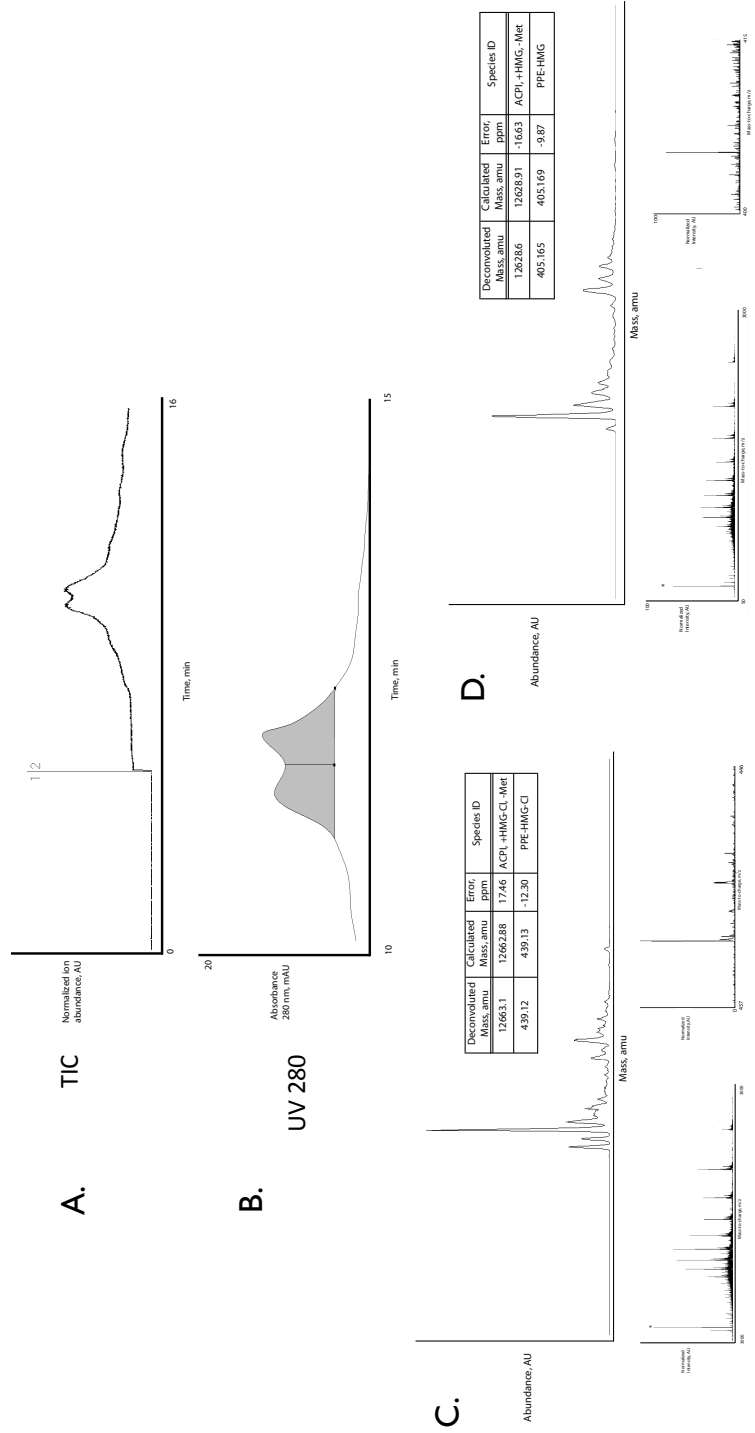


Figure 28. Analysis of ACP-Hal interaction via LC-MS/MS. Representative dataset demonstrating separation of and quantification of chlorinated and nonchlorinated ACP isoform. A. Total ion chromatogram. B. UV 280nm chromatogram. C. Deconvoluted mass of HMG-ACP, inset full charge distribution and phosphopantetheine ejection ion. D. Deconvoluted mass of 4-chloro-HMG-ACP, inset full charge distribution and phosphopantetheine ejection ion.

2.2.6 Interrogation of Hal-ACP interaction by computational docking

To further our understanding of the ACP-Hal interactions a docked model of the proteins was prepared using the online server Hex (54). The crystal structure of the closed form-halogenase containing the Cl⁻, Fe²⁺ and α -KG cofactors (PDB- 3NNF) (55), was used as the receptor ligand, while the NMR structure of the HMG loaded-ACP₁ (PDB- 2LIW) (53) was used as the ligand protein. Residues identified from the ACP activity used to guide the investigation of ACP-Hal interactions that are effected by mutagenesis.

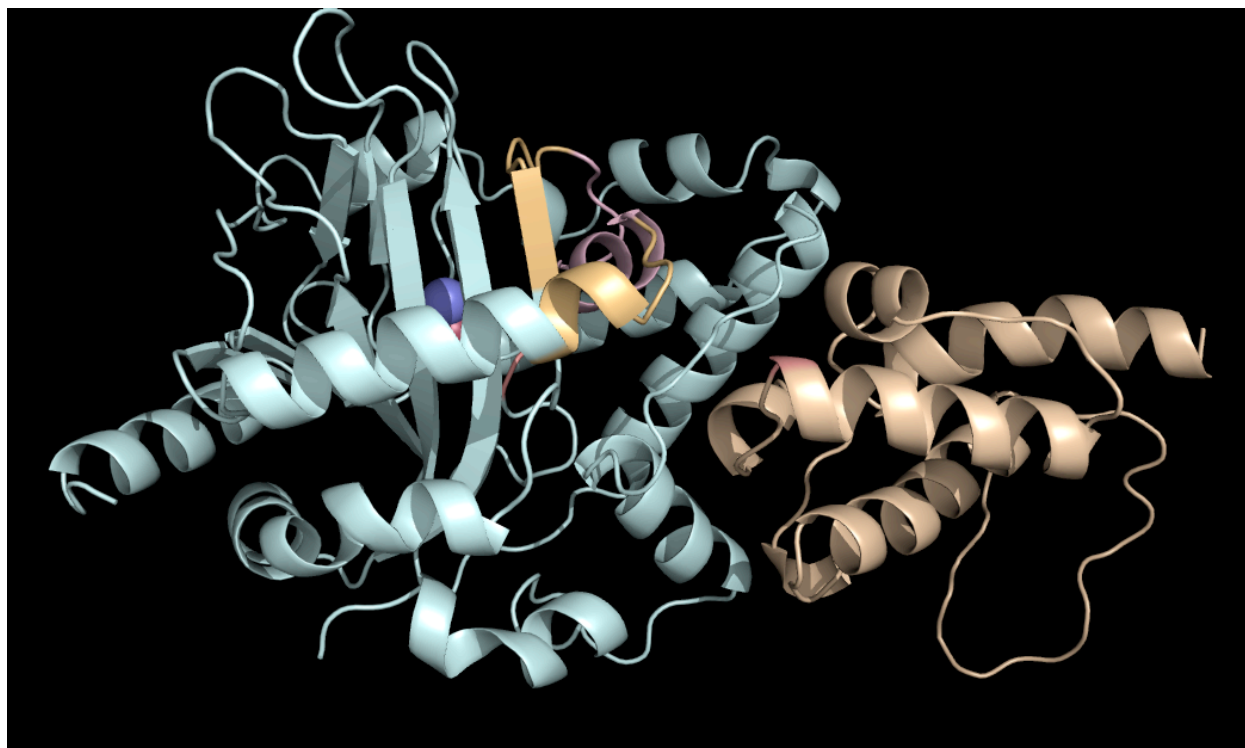


Figure 29. Computationally docked model for CurA Hal and ACP. Cyan- Hal. Pink- Active site lid. Blue sphere- active site iron. Red sphere- active site chlorine. Wheat- ACP, Salmon- active site serine. See text for details.

The ACP active site serine, on the interhelical loop between α 1- α 2, is angled directly into the Hal active site, sitting below the Hal lid (Figure 29). The distance from the serine to the chlorine density is approximately 25Å, within reason for a fully extended HMG-PPant to interact in the Hal active site. Interestingly Hal R279 sits directly over the ACP Ser1999 which likely functions to stabilize, via charge complementation, the PPant phosphate (Figure 30).

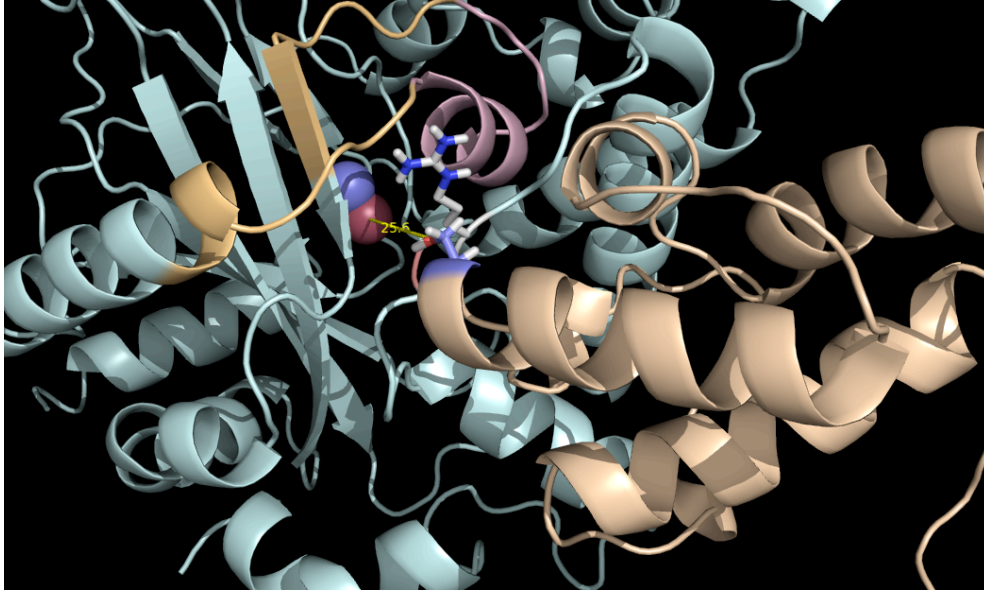


Figure 30. ACP-Hal active site interaction. CurA ACP projects into the Hal active site. Hal R279 is positioned over the ACP active site likely interacting with the ACP prosthetic group.

Extensive salt bridging interactions and hydrogen bonding are observable between the docked proteins. Readily apparent is the direct interaction of ACP D1988 and Hal K282 (Figure 31), as well H-bonds formed between the conserved ACP YDY2013-2015 on the C-terminus of ACP α 3 forming with side chains from Hal Q45, K275, and R272 (Figure 32).

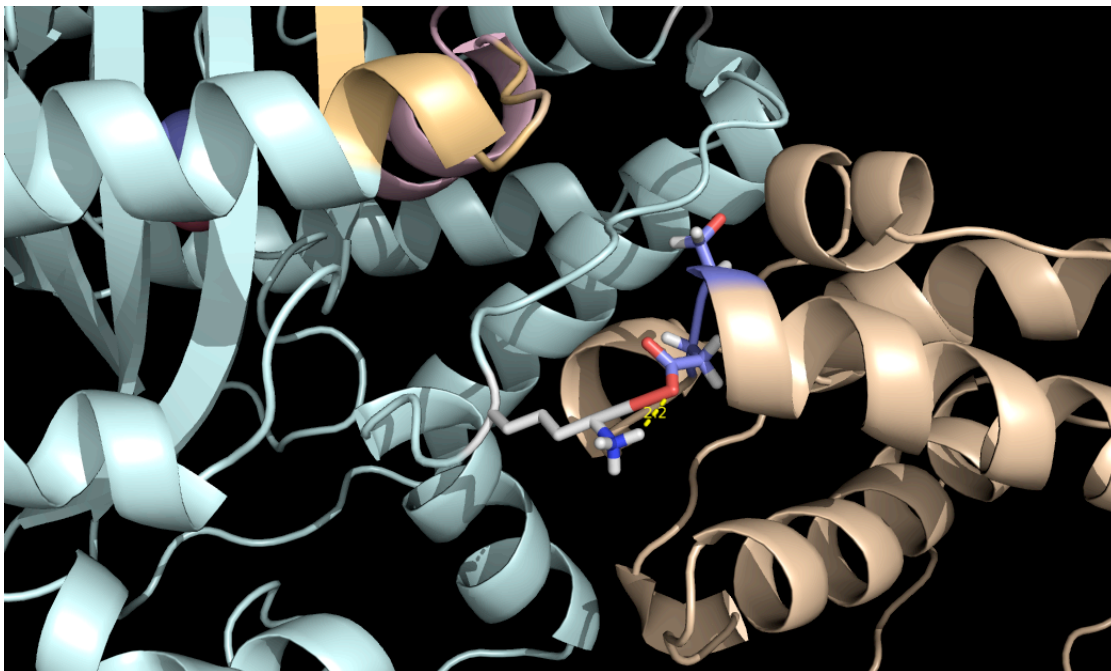


Figure 31. Salt bridge interactions are disturbed upon ACP D1988 mutation.

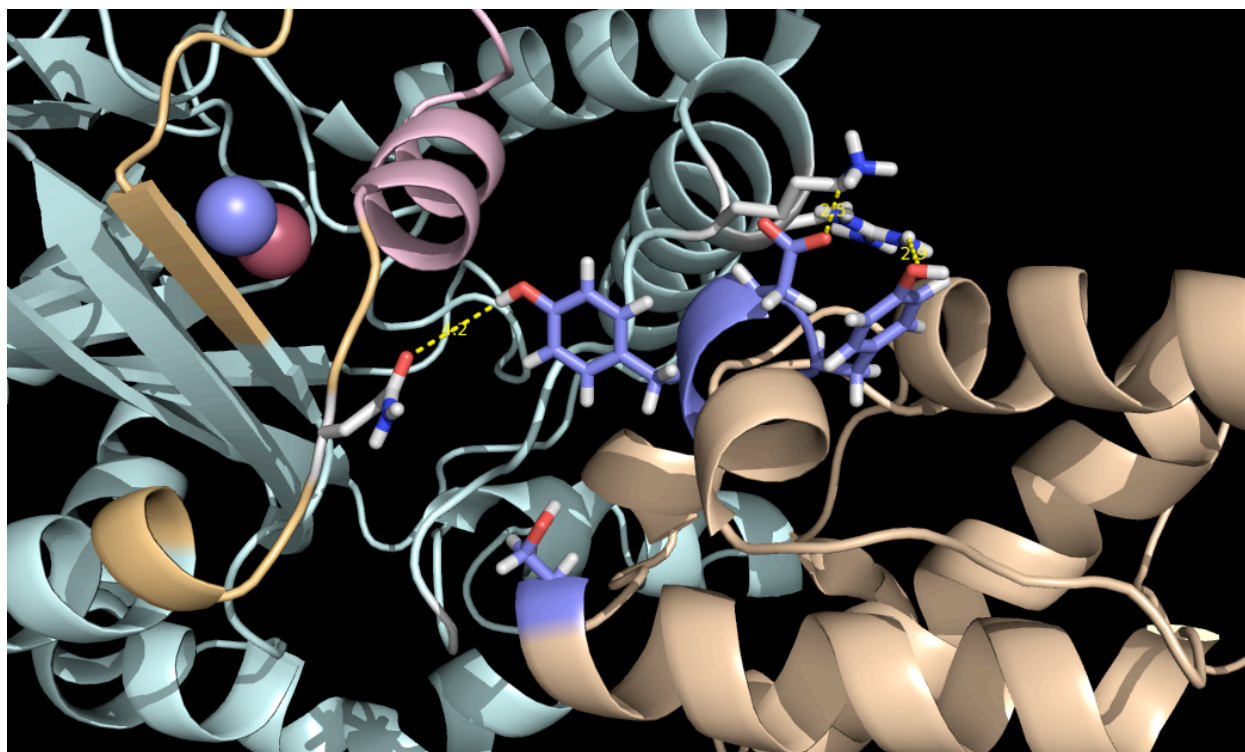


Figure 32. ACP hydrogen bonding is disrupted in site directed mutants. Highly conserved Y-D-Y, residues 2013-15 engage the Hal in multiple surfaces interactions.

The effect of mutations on ACP $\alpha 2$ and the N-Terminus of ACP $\alpha 3$ are likely the results of perturbations to the densely packed hydrophobic network (Figure 33). Insertion of the large bulky side chain as in the A2009R mutant also likely effects the presentation of the substrate to the halogenase. It is apparent, that surface contacts between CurA ACPs active site and Hal play an important role for inter-domain recognition and catalysis, though a discrete function and the specific contacts enabled by Cd remains elusive.

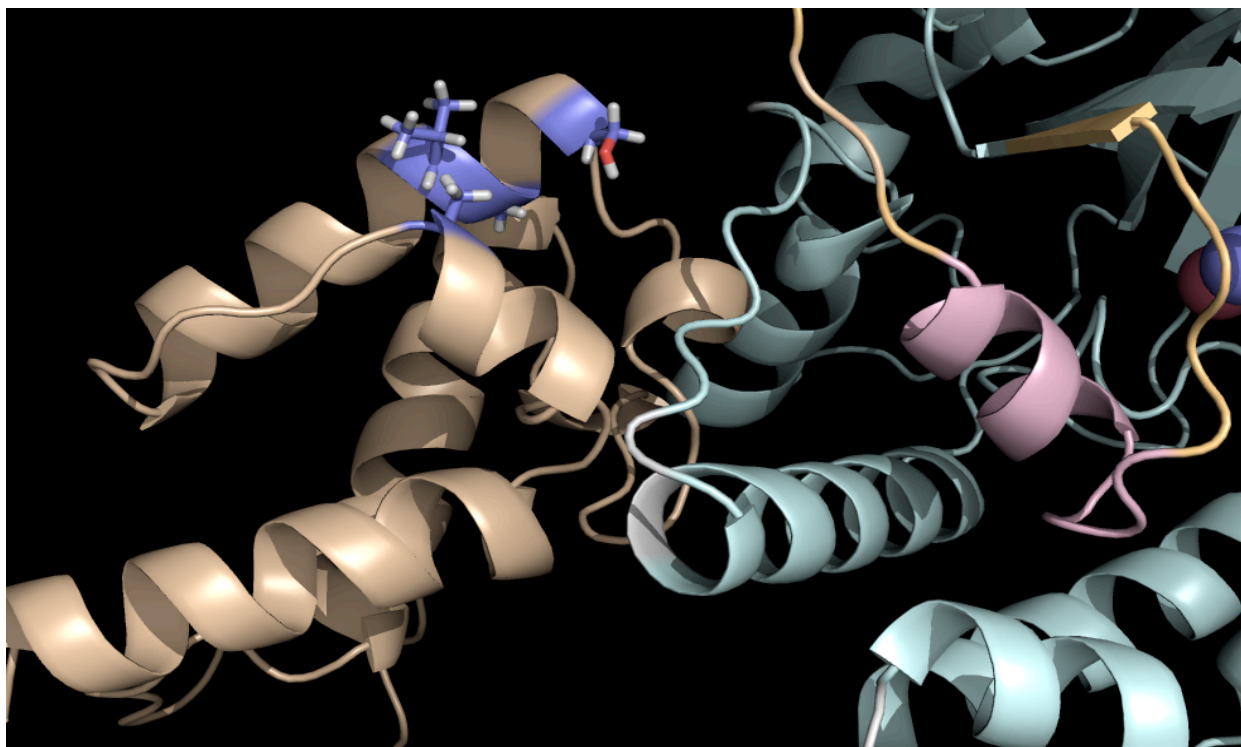


Figure 33. Tightly packed hydrophobic network on the ACP is disrupted by mutation.

2.2.7 Putative structure of the Cd resembles PKS docking domains

As a complement to the docking study, the putative structure of the Cd was determined by ITASSER structural prediction. Comparison of the Cd, with PKS dd and NRPS com show the high degree of structural similarity of the Cd with the helix-loop-helix architecture of the erythromycin docking domains (50, 56), as contrasted with the communication domain from tubulysin pathway (52).

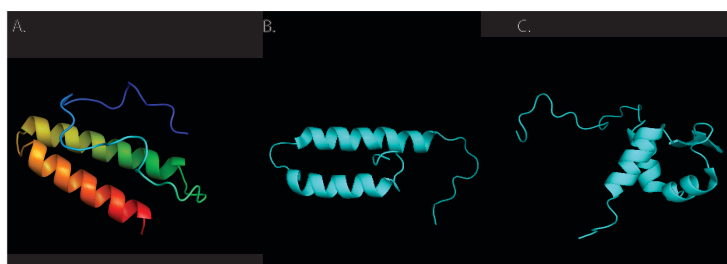


Figure 34. Structural comparison of donor modules from intermodular system. A. CurA Cd. B. PKS-DEBS2 docking domain. C. NRPS- TubC communication domain.

2.3 Discussion

While β -branching systems are not common, they are by no means rare in natural product biosynthesis. The six core HCS cassette proteins also appear as biosynthetic components of other NPs such as virginamycin (*Streptomyces*), myxovirescin A (*Myxococcus*), bacillaene (*Bacillus*), bryostatin and mupirocin (*Pseudomonas*), each resulting in a distinct structural moiety. The Cur and Jam systems are unique in that inclusion of an additional catalytic domain, Hal, is critical for generating either the 2-methyl-cyclopropyl group or vinyl chloride, respectively. Additionally, the presence of C-terminal ACP flanking domain, Cd, conspicuously present exclusively in the Hal containing pathways drives the question of whether these domains are engaging in specific intramodular protein-interactions as a means of passing substrate back up-stream following modification by the HMG-CoA synthase.

As demonstrated during the NMR titration experiments, individual ACP domains do not interact. This finding would portend of independently functioning carrier proteins domains as the full-length triplet. Consequently, presence of a somewhat large portion of dimeric ACP triplet, lacking the Cd domain observed by EM is perplexing. The presence of dimeric proteins can potentially be explained by considering the conditions under which the ACPs were prepared. Using the previously published protocols for overexpressing ACPs, that is in unbuffered media lacking Fe^{2+} and an overexpression time of greater than six hours, undoubtedly a fraction of ACP phosphopantetheinylated. Given the absence of reducing agent from the purification buffers, and the significant amount of time proteins are exposed to air during prior to EM analysis, it stands to reason that a significant portion of PPant disulfide could easily have formed. Additionally, the construct of ACP_{I+II+III} that was used for EM analysis possessed a serine to alanine mutation for the ACP_{II}, thus one third of the potential active site modifications could not occur and this may account for the decrease in disulfide formation.

The elongated structure of the protein, as determined by AUC measurements might also be useful in understanding the unusual SEC profile of the ACPs. Molecular weight calculations by Gel-filtration or Gel-permeation chromatography are predicated based on the construction of a standard molecular weight plot. For macromolecules, proteins or protein complexes, are assumed to run as globular, sphere-like proteins. Indeed, analytical size exclusion

chromatography performed on the tandem ACP-array from PUFA biosynthesis, which has no Cd-like domain, demonstrated a similar, greater-than-expected molecular weight elution profile; excised carrier proteins were shown to elute at 1.69-2.68-fold greater than expected, while the pentadomain eluted at a volume equal to 2.98-fold greater molecular weight (57). Moreover, the small angle X-ray scattering (SAXS) derived mode for the tandem ACP, reveals that the fused proteins are overall monomeric and that the flexible interdomain linkers enable autonomous functioning of individual ACPs. The increase apparent hydrodynamic volume, resulting from of the beads-on-a-string conformation, can also be used to explain differences between the expected and observed analytical chromatography profiles for each of the ACPs. This type of domain mobility is also observed under negative stain conditions with the Cur ACP triplet, suggesting that the inherent protein flexibility may also play a role in the functioning of those proteins and the lack of substantial binding affinity as determined by NMR titration and ITC, though validity of significance of the dissociation constant obtained by dilution experiment is questionable given that the fitted data extend outside the tested range.

The strongest support for functional role of Cd comes from a consideration of the stepwise formation of the 2-methyl-cyclopropane. Halogenation rates appeared to be the only reaction during which the Cd demonstrated a significant effect, though minor differences in the proportion of reactant ACP proceeding towards dehydration and decarboxylation were observed. An experiment can be designed in order to dissect the potentially enhanced Cd mediated interaction with ECH₁, catalyzing the reversible dehydration reaction. Carrying out the HMG-ACP, Hal and ECH₁ reaction in H₂¹⁸O and using a highly resolving mass spectrometer (i.e. FT-ICR-MS) the equilibrium between the proteins species could be examined. If the reaction were demonstrated to be reversible, the hydroxyl group on the HMG would carry a 2 Da increase in mass, resulting from the heavy oxygen, while modifications could similarly be localized to the phosphopantetheine bound substrate using the PPant ejection assay. Rates of formation could then be correlated to enhanced interaction for the ACP_{III} or ACP_{III+Cd}. A possible explanation for the lack of observed difference in activity of the WT and N-terminally truncated ECH₂ is the unusual quaternary structure of the excised domain. ECH₂ was demonstrated to be trimeric (58) in solution, rather than the prototypical dimer for PKS proteins. The addition of this third domain may be sufficient to mask any effective increase in the presence of the Cd. The

oligomeric state of the ECH₂ has been shown to be concentration dependent, and thus running the reactions at a lower concentration of decarboxylase may reveal previously undetected differences in conversion.

A direct quantification of the affinity of ACPs for self association and β -branching proteins can be examined using a variety of methods including fluorescence polarization, surface plasmon resonance or biolayer interferometry, which would provide conclusive evidence as to the origins of biosynthetic flux increases associated with the Cd.

2.4 Materials and methods

Bacterial strains, media and culture conditions.

Escherichia coli DH5 α subcloning efficiency cells (Invitrogen) were used as axenic stocks for the preparation of electrocompetent cells for DNA propagation and maintenance. *Escherichia coli* Rosetta BL21(DE3) (Novagen) harboring the pRARE plasmid (spectinomycin resistance exchanged for chloramphenicol by Dr. W. Clay Brown, University of Michigan) for general protein overexpression; holo- carrier proteins were obtained using the BAP1 cell line (Pfeifer).

All transformations were achieved via electroporation (BioRad, Gene Pulser Xcell), with 2.5-500 ng DNA from ligation reaction mixture or as intact plasmid, recovered with SOB, allowed to outgrow for 60 minutes prior to plating on lysogeny broth-Miller-agar plates (LB-Miller) supplemented with the appropriate antibiotic (ampicillin- 50 μ g/mL, kanamycin- 50 μ g/mL, spectinomycin- 25 μ g/mL).

Protein overexpression, media and culture conditions

Fresh transformants harboring the plasmid of interest were used to inoculate a culture of lysogeny broth-Miller (LB-Miller), corresponding with plasmid resistance marker (ampicillin- 50 μ g/mL, kanamycin- 50 μ g/mL, spectinomycin- 25 μ g/mL). Overnight (~12-18 hours) cultures were used to subinoculate (1:100 (v/v)) large-scale (500-1000 mL) cultures supplemented with appropriate antibiotics for protein overexpression. Apo- carrier proteins were cultured in LB-miller supplemented with 1% (v/v) Glycerol, 25mM Tris (pH= 7.5), and freshly prepared 100 μ M Fe(II)SO₄. ECH₁ was grown in 2XYT, while Hal, ECH₂WT, ECH₂dd, ERA Δ 2 and Sfp were all grown in LB-Miller.

For all proteins of interests, large-scales cultures were grown at 37° while shaking at 160 RPM until reaching an OD₆₀₀ 0.6-0.8. Flasks were allowed to cool at room temperature for 20-30 minutes. Protein overexpression was induced with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM. Flasks were returned to a 22° shaker and continued to grow to lower the temperature of the incubator by 2 degree increments over 20 minute intervals until a final temperature of 18° was reached. Carrier proteins were harvested following 5 hours induction, while Hal, ECH₁, ECH₂WT, ECH₂dd, ERA₂ and Sfp were allowed to grow for 12-16 hours prior to harvesting. Cells were harvested by centrifugation at (5000xg, 10 minutes, 4°), transferred to 50mL Falcon tubes and cell pellets were flash frozen in liquid nitrogen and stored at -80° until purification.

Protein purification

All steps during the protein purifications were performed at 4°. Cell pellets were resuspended in 2x cell volume (minimum 25mL) 5-10% Buffer B, and incubated with freshly prepared lysozyme (0.2 mg/mL final, prepared in MilliQ H₂O), freshly prepared PMSF (1 mM final, prepared in DMSO) and 0.5 *cOmplete* protease inhibitor cocktail (Roche) tablet for 30 minutes while rocking. Cells were lysed by use of a recently tuned ultrasonic dismembrator with 5 cycles of 2 minutes, with pulses of 3 seconds on, followed by 3 seconds off, with intensities ranging from 7-8. All cycles were carried with chilled beakers in packed ice bucket. Cellular debris was removed by centrifugation at 40000-50000xg for 30 minutes and the supernatant was transferred to a fresh tube.; residual insoluble material was removed by filtration (syringe driven 0.45 μm filter) or a second round of centrifugation.

Immobilized metal affinity chromatography

The first step to purifying all proteins made use of the affinity of the His-tag towards divalent cations. Clarified lysate was applied to 2 mL nitrolotriactic acid-agarose resin (Qiagen), charged with Ni²⁺, and equilibrated with 10% Buffer B. The resin was washed extensively with 25 column volumes of 15% buffer B to remove any nonspecifically bound proteins. The target protein was eluted with a stepwise gradient from 20-100% Buffer B. Peak fractions as determined by gradient SDS-PAGE ((4-12%) (Novex NuPage) run in MES buffer at 180 mV for

50 minutes) were pooled and concentrated using Amicon ultra centrifugal concentrators (Millipore, 0.5 kDa, 3 kDa, 10 kDa), equilibrated with Buffer F until a volume of 2.5 mL was reached. Proteins were buffer exchanged with a PD10 washed and equilibrated with Buffer F and eluted with 3.5 mL of Buffer F. Proteins were further concentrated to less than 100-500 μ L, aliquoted, flash frozen in liquid nitrogen and stored at -80° .

Analytical Gel Filtration

Proteins used for FP, ITC and EM analysis were further purified by size exclusion chromatography. Concentrated aliquots were loaded onto a Superdex 200 10/300 GL (GE Healthcare) equilibrated with either Buffer F (FP) or Buffer G (ITC, EM). Fractions were analyzed for purity by SDS-PAGE, concentrated, aliquoted flash frozen in liquid nitrogen and stored at -80° for FP and EM; proteins used for ITC were dialyzed (1:4000) against freshly prepared Buffer G and subsequently concentrated and stored. Protein concentrations were determined using the bicinchoninic acid assay (BCA Pierce) according to the manufacturers guidelines, using a BSA standard curve and read in 96 well format.

Analytical Ultracentrifugation.

ACP3 and ACP3-CD were analyzed by sedimentation velocity in 50 mM Tris, 200 mM NaCl, 5% Glycerol at 0.2, 0.5 and 0.8 mg ml⁻¹ using a ProteomeLab XL-I (Beckman Coulter). Samples were loaded into sector-shaped double channel centerpieces and spun at 42,000 rpm in an AN50TI rotor and 22 $^{\circ}$ C. Absorbance scans were collected continuously at 280 nm and a radial resolution of 40 μ m and 3 replicates. Data analysis was with the enhanced van Holde-Weischet analysis module followed by 2-dimensional sedimentation spectrum analysis (2-DSA) using the finite element modeling module provided in the Ultrascan software (<http://www.ultrascan.uthscsa.edu>). Sedimentation profiles were analyzed at a grid resolution of 25600 using 16 grid repetitions. Confidence levels were derived from 2-DSA data refinement using genetic algorithm followed by 50 Monte Carlo simulations. Calculations to analyze analytical ultracentrifugation data were performed on the UltraScan LIMS cluster at the Bioinformatics Core Facility at the University of Texas Health Science Center at San Antonio, the Lonestar cluster at the Texas Advanced Computing Center (supported by NSF Teragrid Grant #MCB070038 to Borries Demeler). v_{Bar} of 0.739 ml/g for ACP3 and ACP3-CD was

calculated from the polypeptide sequence and the buffer density (1.0235 g/ ml) and viscosity (1.2116 cp) were calculated using the buffer calculation module provided with the Ultrascan software suite.

Dilution experiments by Isothermal titration calorimetry

Dialyzed proteins prepared from gel filtration fraction were concentrated to the stability limits of the protein. Using a MicroCal ITC, 300 μ L of sample were loaded into the syringe and 1.7 mL of the dialysate was added to both the sample and control wells. 25 total injections, 5 μ L for first injections and 10 μ L for the remainder, were made into the sample cell held at 25°, with a spacing of 150 seconds and a stir speed of 310 rpm. Data were processed using the Origin 7 software suite according to the manufacturers guidelines and fit to the dilution function.

Mass spectrometric analysis of proteins by LC-ESI-QTOF-MS

The identities of the proteins in this study as well as the extent of phosphopantetheinylation of ACPs were determined by electrospray-time-of-flight mass spectrometry. Protein samples were removed from the concentrated fraction prior to buffer exchange (minimum 2 mg/ml), diluted 1:10, acidified with 10% (v/v) formic acid (FA), centrifuged at 20800xg for 10 minutes and transferred to a fresh tube. Of this sample, 20 μ L was injected onto a C4 column (Phenomenex Aeris™ 3.6 μ m WIDEPOR C4 300 Å, LC Column 50 x 2.1 mm) equilibrated with 30% CH₃CN+0.1%FA, connected online with an Agilent 6250 quadrupole – Time-of-Flight mass spectrometer. Column elution was diverted to waste for sample desalting for 1 minutes prior to a gradient of 30-95% CH₃CN+0.1%FA over 8 minutes, followed by a 1 minute wash at 95% CH₃CN+0.1%FA and reequilibration at 30% CH₃CN+0.1%FA for 2 minutes. For MS experiments requiring separation of halogenated and nonhalogenated ACPs, elution from a C8 column (Agilent Zorbax 300SB 3.5 μ m, 50 x 4.6 mm) equilibrated with 30% CH₃CN+0.1%FA+0.01% TFA was diverted to waste for sample desalting for 1 minutes prior to a gradient of 30-60% CH₃CN+0.1% FA+ 0.01% TFA over 35 minutes, followed by a 5 minute wash at 95% CH₃CN+0.1% FA+ 0.01% TFA and reequilibration at 30% CH₃CN+0.1%FA+0.01%TFA for 10 minutes.

All spectra were processed using the Bioconfirm Intact Protein module on the Agilent MassHunter Qualitative Analysis software. Protein identities were confirmed by obtaining an averaged mass spectrum over maximum peak height as determined by Total Ion Chromatogram (TIC) with a 30second background spectrum subtracted. Spectral deconvolution was performed over a range of 600-3000 m/z, using a minimum of 5 consecutive charge states and fit score of 7, with a limited mass range of $\pm 15\%$ of the calculated mass of the protein.

Preparation of HMG-loaded ACP

Apo- carrier proteins, as determined by mass spectrometry, were loaded with racemic 3-hydroxy-3-methylglutaryl Coenzyme A (HMG-CoA). In a 1000 μL reaction (50 mM Tris (pH= 8.0) 50 mM NaCl, 1 mM MgCl_2 , 10% (v/v) glycerol, 1 mM TCEP)), 150-300 μM ACP was incubated along with 5 μM recombinant *Bacillus subtilis* phosphopantetheinyl transferase, Sfp. The reaction was initiated with the addition of freshly prepared HMG-CoA prepared as a 5 mM stock in MilliQ H_2O . 5, 7.5 and 10 molar equivalents of HMG-CoA were used to ensure complete loading for single excised ACPs, ACP didomain and ACP tridomain constructs, respectively. Reactions were incubated at room temperature for 2 hours followed by dilution to 1.5mL with Buffer F. The diluted reaction was loaded onto a HiTrap Desalting column (GE Healthcare) equilibrated with Buffer F and eluted with 2mL Buffer F, to remove unreacted HMG-CoA, TCEP and salts. Proteins were concentrated, aliquoted, flash frozen in liquid nitrogen and stored at -80° . The concentration was measured and extent of substrate loading was determined by mass spectrometry as above.

BODIPY Labeling of apo-GCCT proteins

Fluorophore bioconjugated ACPs were prepared from apo- carrier proteins harboring engineered cysteines, introduced by site directed mutagenesis and labeled with thiol reactive BODIPY N-(4,4-Difluoro-5,7-Dimethyl-4-Bora-3a,4a-Diaza-s-Indacene-3-yl)Methyl)Iodoacetamide) (Invitrogen). Briefly, in a 1000 μL reaction, 50-75 μM ACP was incubated in Buffer F with 5 equivalents of TCEP at room temperature for 10 minutes. BODIPY-FI-C1-IA was prepared fresh in DMSO as a 10mM stock and added dropwise to the reaction in 5-10 molar excess of the ACP. The reaction was allowed to proceed in the dark at room temperature for 2 hours. The reaction was quenched with 100 equivalents of β -mercaptoethanol followed by dilution to 1.5 mL with

buffer F. The diluted reaction was loaded onto a HiTrap Desalting column (GE Healthcare) equilibrated with Buffer F and eluted with 2 mL Buffer F, to remove unreacted BODIPY, TCEP and β -mercaptoethanol. Proteins were concentrated, aliquoted, flash frozen in liquid nitrogen and stored at -80° . The concentration was measured and extent of labeling was determined by BCA assay and mass spectrometry as above.

Halogenase activity assay

Enzymatic assays were performed in order to analyze the activity of each of the β -branching proteins successively. For each enzyme a 220 μ L reaction (final volume), 50 mM Tris (pH= 8.0), 50 mM NaCl, 10% glycerol, 1mM α -ketoglutaric acid (pH= 8.0) and freshly prepared 100 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ were mixed. 5 μ M halogenase was added to the reaction mixture and allowed to incubate at room temperature for 5 minutes. Assays were initiated with the addition of HMG-ACP, prepared as above. 25 μ L reactions aliquots were removed at specific time points and quenched with 10% formic acid. Reactions were stored at -80° until processed. All reactions were performed in triplicate.

Quenched reactions were loaded onto a Pierce micro spin column packed with 40 μ g C4 resin (ANALTECH) equilibrated with MilliQ $\text{H}_2\text{O}+0.1\%$ FA and dried. Samples were washed twice with 120 μ L MilliQ $\text{H}_2\text{O}+0.1\%$ FA, and eluted with 50 μ L 100% $\text{CH}_3\text{CN}+0.1\%$ FA. Reactions were analyzed by mass spectrometry as above.

Relative activities of individual β -branching enzymes were quantified by comparing the abundances of the deconvoluted ACP masses corresponding to all transformation under investigation.

Dehydratase, Decarboxylase and Cyclization reaction monitoring.

Reactions of the ECH₁, ECH₂WT, ECH₂dd and ER, were prepared as described for the Halogenase assay. Assays were allowed to incubate at room temperature for 30-45 minutes in order generate the maximum amount of the requisite ACP bound substrate. 5 μ M of respective

enzymes were added and 25 μ L reactions aliquots were removed at specific time points and quenched with 10% formic acid. Reactions were processed as above.

One-pot assay for cyclopropane formation

Assay mixtures were prepared as described for the Halogenase activity monitoring along with the Halogenase, ECH₁, ECH₂WT, and ER. The reactions were allowed to incubate at room temperature for 10 minutes prior to the addition of the HMG-ACP. Time points were removed, worked up and analyzed as described above.

Acknowledgements: This work was performed collaboratively between the Sherman, Skiniotis, Walter and Dötsch and Smith labs. Dr. Liangcai Gu (LG), Eli Benchell Eisman (EBE) and Dr. Alena Busche (AB) generated all ACP constructs used in these studies; Hal, JamJ ECH₁, ECH₂WT were generated by LG, ERA2 was prepared by Dr. Dheeraj Khar, ECH₂dd and Cd were generated by EBE. One-pot assays for small molecule production were conducted and analyzed by LG. All protein LC-MS experiments were conducted by EBE. Protein preparations including analytical gel filtration for the AUC and EM were conducted by EBE. AUC experiments were done by Dr. Titus Franzmann and Prof. Stefan Walter. EM experiments were performed by Dr. Somnath Dutta and Prof. Yiorgio Skiniotis. AB and Prof. Volker Dötsch, performed all NMR experiments and solved the solution structure of the excised CurA ACP. Computational docking was performed by EBE.

Chapter 3

Activated-elimination as a strategy for hydrocarbon production

3.1 Introduction

The adoption of biofuels as a renewable energy alternative has been slow as a consequence of the high cost and low processing efficiency of current methods in the acquisition of high-value, fuel-like molecules. Additionally, the absence of an infrastructure to support large-scale implementation of biomass-derived alternatives has led to a protracted shift in the energy paradigm. This has led to a renewed impetus for the identification of drop-in replacements for fossil fuel and fostered the development of second-generation biofuels. These novel surrogates are capable of being attained from non-food related feedstocks as well as microbiological sources. In order to address the growing need for a sustainable energy future, we have begun to harness the potential of a simple enzyme system from the marine cyanobacterium *Lyngbya majuscula*, demonstrating the capacity to produce molecules chemically identical to fossil fuels *in vitro*, working towards the establishment of a novel advanced biofuels platform.

Our previous investigation of the curacin A bioassembly revealed a unique mechanism by which the natural product is released as the terminal alkene (Figure 35) (36). Briefly, malonyl-CoA is activated by the CurM acyl transferase (AT) domain and transthioesterified to the CurM acyl carrier protein (ACP). The CurM ketosynthase (KS) transfers the biosynthetic intermediate from the preceding module, decarboxylates the ACP bound malonate and catalyzes a Claisen-like condensation extending the growing chain by an acetate unit. The β -keto group is then reduced to the alcohol with (*R*) stereochemistry by the keto reductase (KR) domain. The sulfotransferase (ST) installs a sulfonate moiety on a newly formed β -hydroxy group, which is followed by

decarboxylation catalyzed by the thioesterase (TE) domain, yielding a terminal olefin. The tandem sulfonation-decarboxylation activity of this enzyme pair has been successfully reconstituted *in vitro* and characterized using a hydrocarbon-like substrate mimic (Figure 35). The focus of the ensuing research seeks to expand on these preliminary findings to harness the potential of cyanobacterial enzymes for the direct production of liquid fuel *in vivo*.

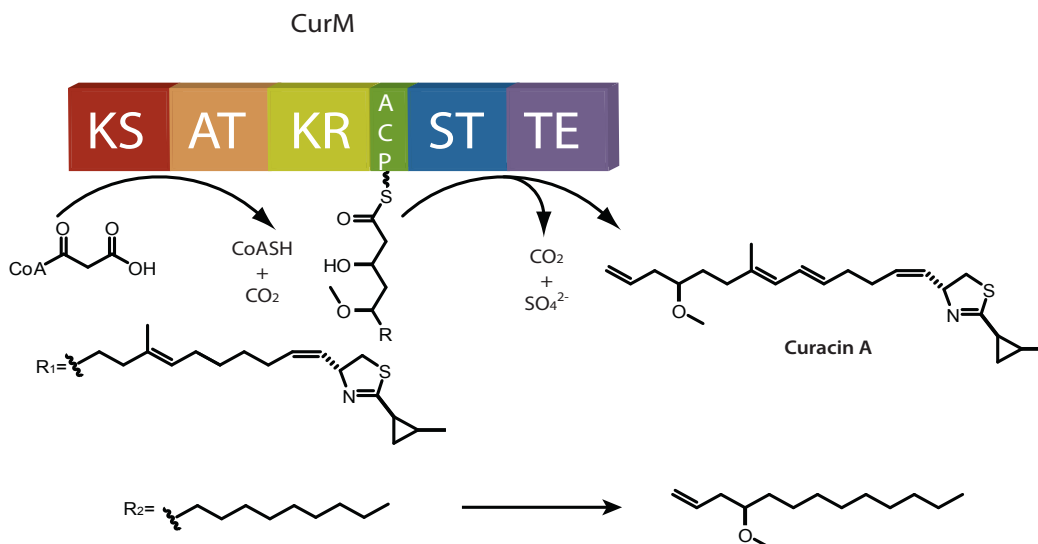


Figure 35. Offloading of curacin A desulfonative-decarboxylation. R1- wild type curacin A. R2- substrate analog used to probe ST and TE functionalities.

3.2 Results

3.2.1 Sulfotransferase-Thioesterase containing genes from additional cyanobacteria

As an extension of our previous work (36) elucidating the desulfonative-decarboxylation mechanism leading to α -olefin formation in curacin A (Cur-) biosynthesis we identified three additional cyanobacterial genes containing the unusual termination sulfotransferase-thioesterase (ST-TE) didomain cassette suggesting that these modules catalyzed a similar offloading of substrate (Figure 36).

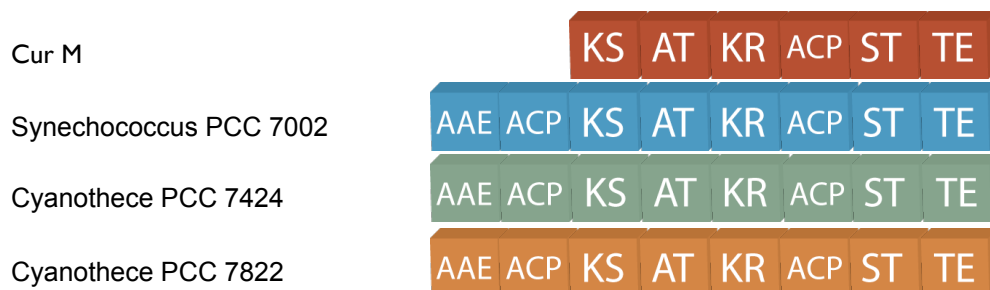


Figure 36. Standalone cyanobacterial pathways identified to contain CurM-like ST-TE domains.

Unlike CurM, which is a component of a larger protein assembly line, these new ST-TE containing pathways appeared as stand-alone modules within the cyanobacterial genomes. The overall sequence homology of the four proteins is 51.1%.

The *in vivo* activity of the ST-TE gene from the *Synechococcus* PCC 7002 was recently characterized by Mendez-Perez and coworkers (59). Through gene deletion studies, the Syn olefin synthase (Ols) was demonstrated to produce 1-nonadecene. The odd-numbered, long-chain alkene is presumed to arise from loading of octadecanoic acid directly on the Syn ACP1 by the acyl activating enzyme, followed by CurM like elongation, reduction and activated-elimination catalyzed by the ST-TE didomain. Motivated by the unique functionality of the Cur and Syn sulfotransferase and thioesterase enzyme pair in producing long chain hydrocarbons, we sought to assess the activities of the homologous domains from the Syn pathway to produce terminally olefinated products *in vitro*.

3.2.1 SynTE is capable of accepting Cur substrate but does not yield alkene

Our initial efforts were focused on the *Synechococcus* PCC 7002 (Syn) thioesterase (TE) and the ability of the protein to catalyze alkene formation. A synthetic construct of the Syn ACP-ST-TE portion of the gene, codon optimized for protein expression in *E. coli*, was obtained and individual domains were cloned as excised proteins and overexpressed according to standard biochemical practices.

Previously, detection of the alkene was carried out by thermal-desorption-GC-MS, however limited access to this instrument, led us to reconsider the analytical methodology. To ensure the

ability to detect low abundance material, a rapid reaction workup and extraction protocol for GC-MS was developed. As positive controls for method validation, reaction-product-like molecules such as octane, decane, methyl 4-methyl-3-oxo valerate, 3-octynol-1-ol and decanal were prepared in reaction buffer at concentrations corresponding with <1% conversion of the ACP loaded substrate to alkene. The samples were treated under standard reactions conditions, extracted, resuspended in a fixed volume, and analyzed by GC-MS (Figure 37).

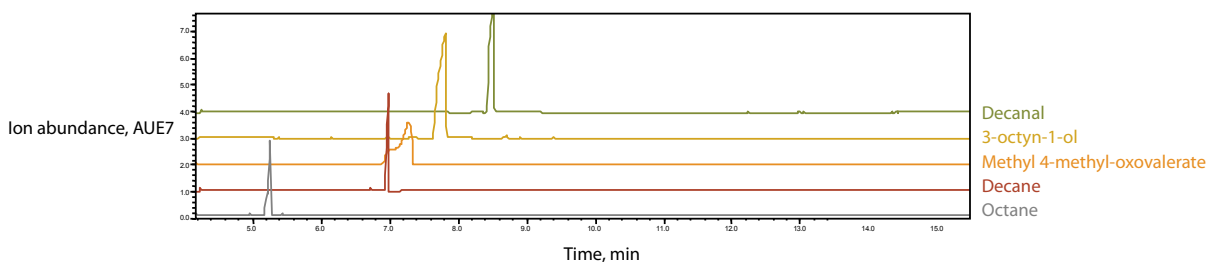


Figure 37. Standards used for GC-MS method development for hydrocarbon detection.

The hydrocarbons were easily separable by GC and gave very abundant signal at concentration below 1000 pg/ μ L by EI-MS, validating both chromatographic and detection methods. Confident in our experimental design for hydrocarbon detection, we proceeded to analyze the ability of the SynTE to accept nonnative substrates and generate olefinated product.

For a standard reaction, 5-(*R*)-methoxy-3-(*R*)-hydroxy-tetradecanoyl-ACP was prepared by incubating apo-carrier protein with the corresponding substrate-CoA and a phosphopantetheinyl transferase from *Streptomyces versicolor*, Svp. Loaded carrier protein was incubated with CurM ST in order to generate the necessary 5-(*R*)-methoxy-3-(*R*)-sulfo-tetradecanoyl-ACP substrate for the TE reactions (Figure 38).

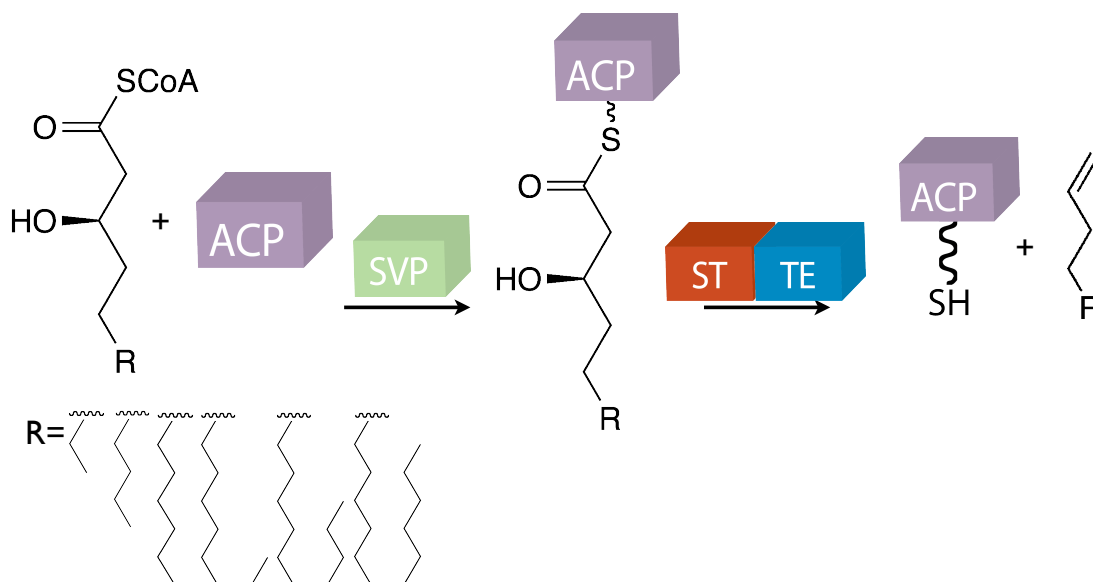


Figure 38. Schematic for sulfotransferase and thioesterase catalyzed reactions.

The assay was initiated following the addition of either CurTE or SynTE Figure. The reactions were quenched, worked up by extraction and analyzed for alkene formation. To our dismay, no hydrocarbon products were detected for either Cur or Syn TE (Figure 39).

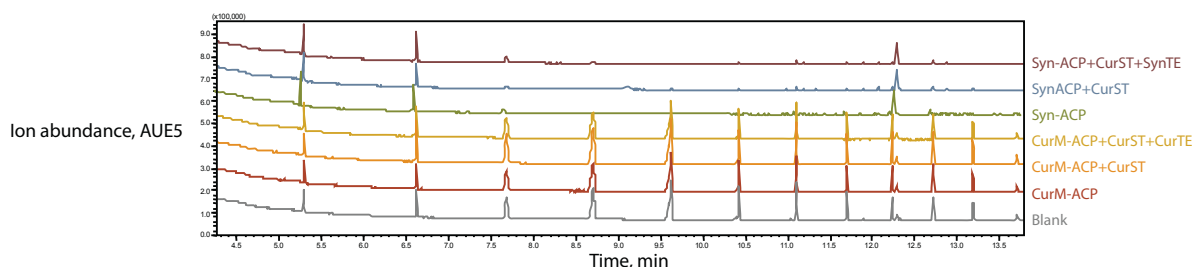


Figure 39. GC-MS analysis of ST-TE reaction workups.

This inability to detect the expected alkene, led us to reexamine the aqueous portion of the reaction for the product of the hydrolysis side-reaction. Proteins were removed from the remaining samples by precipitation and the soluble fractions were analyzed by LC-MS. Indeed, the β -sulfo acid ionized well (Figure 40), suggesting that alkene formation occurs at a substantially lower rate than 1% of reaction turnovers.

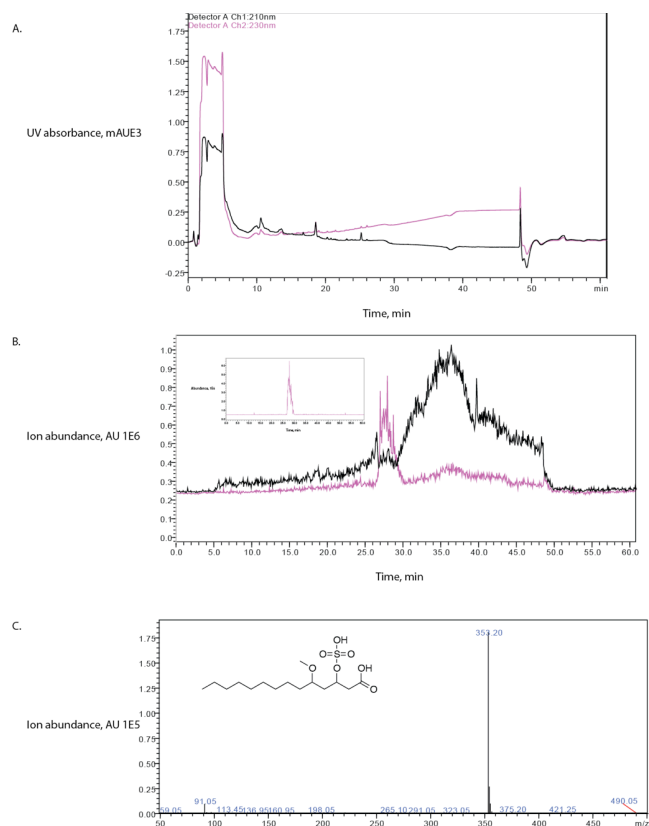


Figure 40. LC-MS analysis of the product of CurST and SynTE. A. UV chromatogram. Black trace- 210 nm. Pink trace- 230 nm. B. Total ion chromatogram. Black trace- positive ion mode. Pink trace- negative mode. Inset- Selected ion monitoring chromatogram $m/z = 353-354$. C. Average mass spectrum 28-28.5 minutes. Expected- 353.36. Observed 353.20

The inability to detect olefin product from either TE reaction prompted a deep investigation of the distinctive function of the Cur and Syn TEs by surveying the unique structure adopted by these decarboxylating thioesterases (**60**). As with other TEs, the Cur enzyme adopt a core α/β hydrolase fold, however significant deviations in quaternary structure from others in this class of protein, provide evidence to account for the uncommon function.

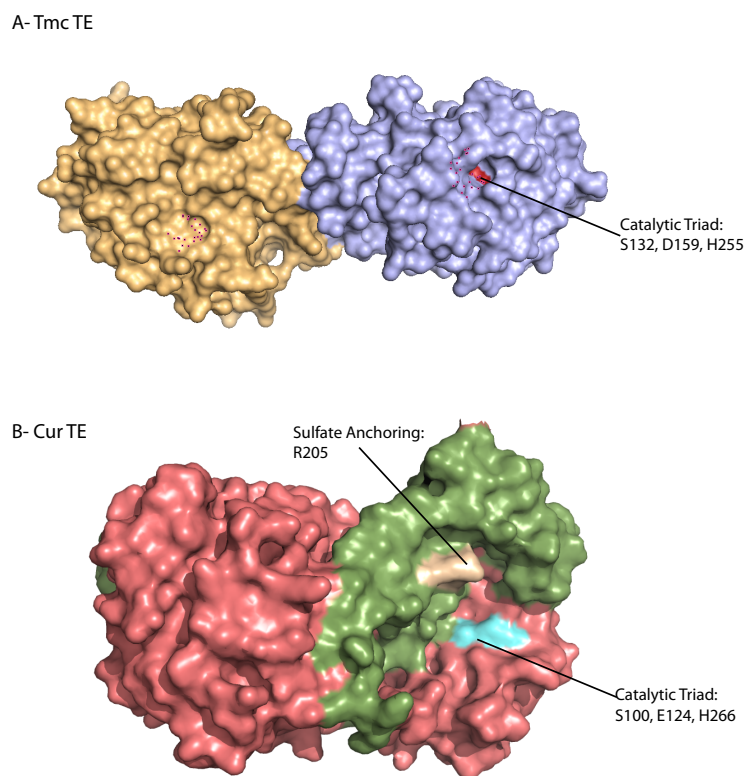


Figure 41. Structural comparison of hydrolysing and decarboxylating thioesterases A. Tautomycin TE. B. Curacin TE.

The unusual core-to-lid dimerization interface of the TE exposes the Ser100-His266-Glu124 catalytic triad, which may promote hydrolytic release, rather than alkene formation (Figure 41). Interestingly, Arg205, an invariant residue in the decarboxylating TEs, is conspicuously present above the active site, likely acting to stabilize interactions with the β -sulfate (**60**, **61**). This result was confirmed whereupon mutant TEs R205Q, R205E, and R205A demonstrated an almost complete loss of conversion of the sulfated substrates; however reactivity towards the β -hydroxy substrates remains unchanged.

3.2.2 SynTE HRSC double mutant prepared for active site labeling protein

To elucidate the origins of the unexpected product discrimination, a SynTE double mutant (SynTE HRSC) was generated, replacing two of the highly conserved catalytic triad Ser-His-Glu residues, S100C and H263R. Identical mutations have previously been utilized for active site

labeling studies characterizing the TE domain from the Streptomyces derived immunosuppressant tautomycin (Tmc) (61).

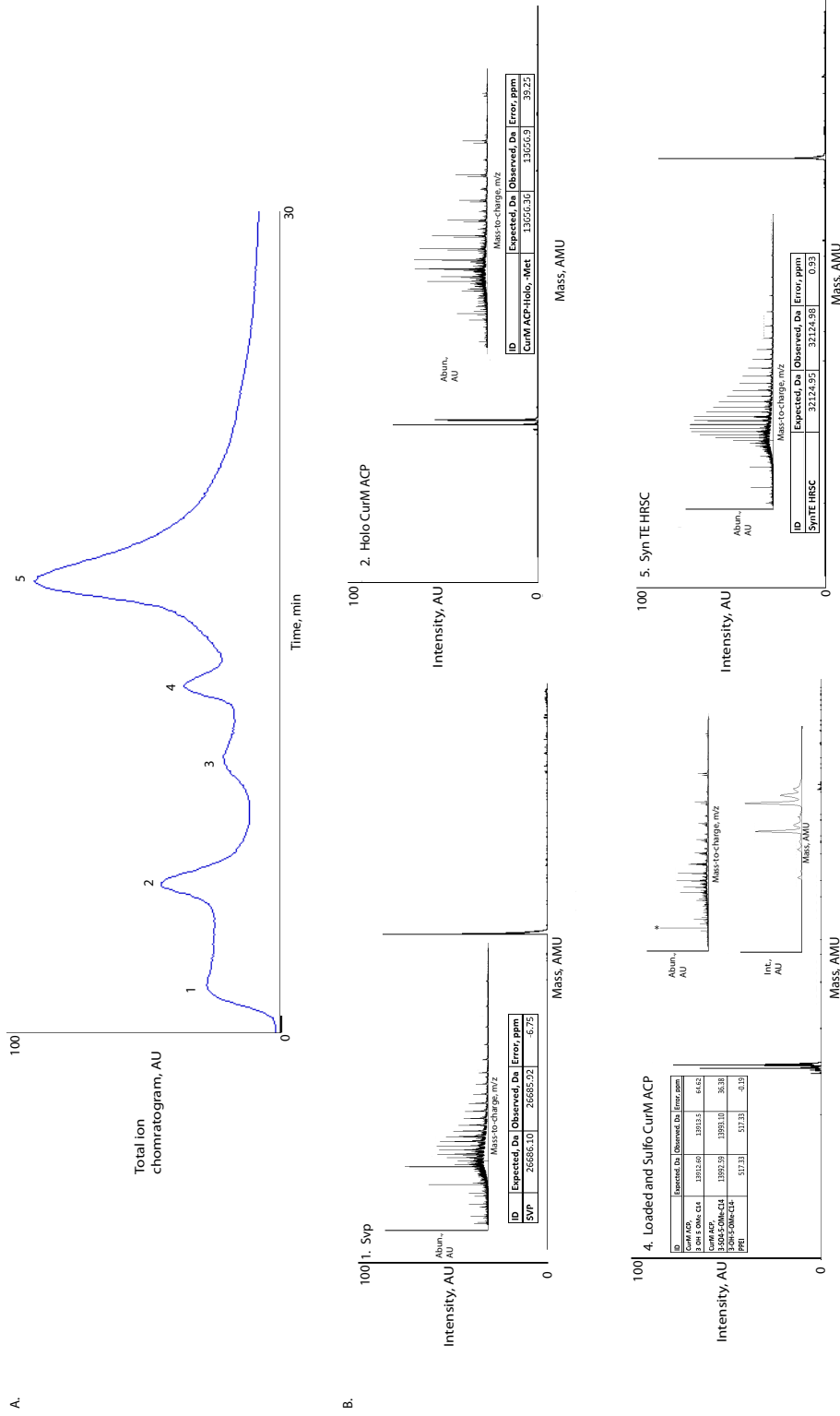


Figure 42. LC-MS analysis of SynTE HRSC labeling. A. Total ion chromatogram and B. charge distribution for individual proteins from acylation reaction as indicated by the corresponding number denoted in TIC; inset: deconvoluted mass 1. Syp, 2. holo-CurM ACP, 3. contaminant protein, 4. coelution of loaded and sulfo-CurM ACP, 5. SynTE HRSC, no detectable labeling.

The activity of the thioesterase was determined by incubating CurM ACP loaded with 3-(*R*)-OH-tetradecanoyl-CoA, CurM ST and either wild type or mutant SynTE and determining a mass balance for the reaction with the hope of capturing substrate in the active site of the TE as a means of determining amino acids implicated in the hydrolysis-alkene formation equilibrium. While the activity of the SynTE HRSC was slightly reduced relative to wild type, as determined by the decreased conversion of the loaded CurM ACP to the holo- form of the protein, no active site labeling was observed using top-down mass spectrometric analysis (Figure 42).

This is likely attributable to the earlier described differences in exposure of the active sites of the TmcTE as contrasted with the SynTE. A feature common to TEs catalyzing hydrolysis is a solvent occluded substrate pocket. However differences in the crystal structures of the Tmc TE and Cur TE clearly demonstrate a novel, solvent exposed active site, with coordination of the sulfate group via positively charged Arg205.

3.2.3 Alternative methods of olefin detection via alkene functionalization

To further characterize olefin production capacity and investigate substrate flexibility of ST-TE containing genes alternative methods of detection of the alkene were explored. Typically, alkyl chains are well suited for measurement by GC-EI-MS, however, as a result of the large percentage of conversion of ACP bound substrate that undergoes the unproductive hydrolysis reaction, rather than alkene formation, direct observation of the low abundance olefin product has been challenging. Additionally, the insolubility of the long chain hydrocarbon under the enzymatic conditions, likely promotes volatilization over the course of the reaction and workup. Taken together this prompted us to look at the use of alkene-enrichment techniques.

Thiol-ene coupling presented itself as a reaction workup method, which would allow for the simultaneous detection and quantification of the sulfo-acid hydrolysis as well as alkene product by LC-MS. Initially pioneered as a solvent-free alternative to thiol-conjugation, thiol-ene chemistry takes advantage of UV stimulated radical formation, mediated through a radical initiator, and leads to the anti-Markovnikov addition of a thiol across a double bond (**62-64**).

Experiments coupling a 1-tridecene standard and the N-acetylcysteamine were successful under typical reaction conditions, including 10% methanol as a solubilizing agent for the radical initiator, benzophenone, generating a product that is amenable to detection by LC-MS (Figure 43). Unfortunately, when the reaction was run, UV exposure caused proteins to rapidly precipitate and no derivatized alkene was detected.

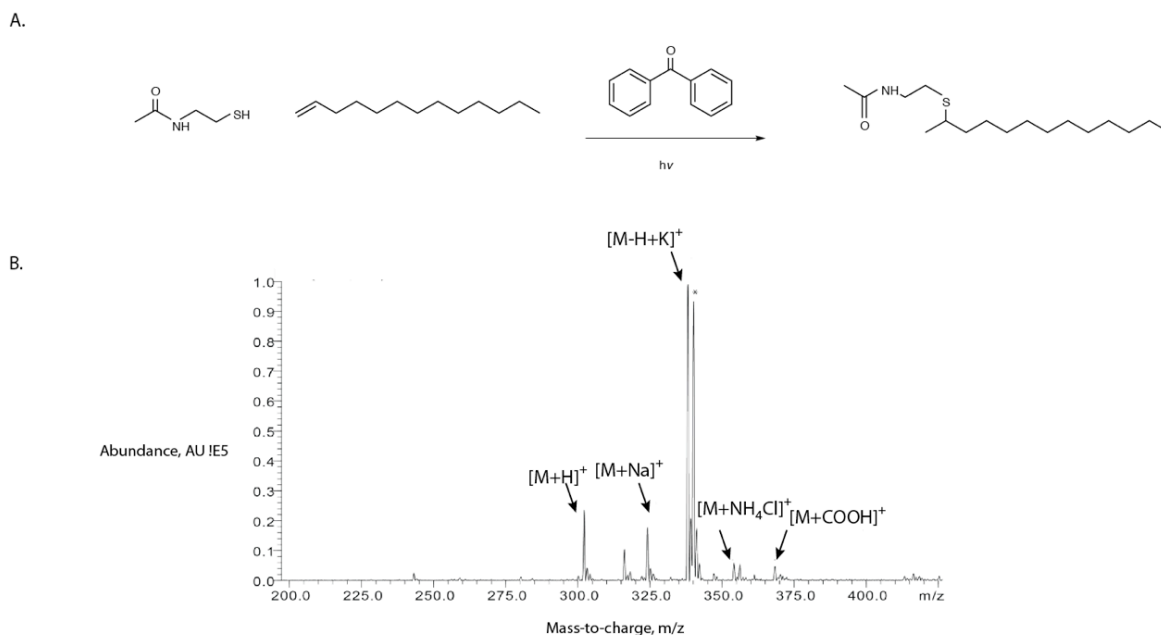


Figure 43. Thiolclick enrichment of alkene containing compounds. A. Scheme for thiol-ene coupling of 1-tridecene and N-acetylcysteamine using benzophenone radical initiator. B. Average mass spectrum of product of thiolclick reaction. Expected $[M+H]$ - 302.24. Observed- 302.29.

3.2.4 Syn and Cur proteins act functionally *in trans* and accept nonnative substrates

The ultimate and penultimate domains from the final module in the curacin biosynthetic pathway constitute a dienzyme cassette that represents a novel PKS offloading mechanism. In our previous investigation into this unusual enzyme pair, we utilized an analog of the terminal curacin chain elongation intermediate, 5-(*R*)-methoxy-3-(*R*)-hydroxy-tetradecanoyl-CoA, which was loaded onto CurM ACP as substrate. This flexibility suggests the enzymes possess reactivity toward related acyl chain analogues containing a β -hydroxy. In order to further probe the promiscuity of the protein pair towards nonnative substrates, and evaluate the production capacity of the CurM enzymes to generate fuel-like hydrocarbons directly, we examined the ability of the sulfotransferase to accept further simplified acyl-substrates.

Synthetic substrates lacking the 5-methoxy group were designed and synthesized in collaboration with the Wipf group at the University of Pittsburgh. 3-(*R*)-hydroxy-tetradecanoyl-CoA was prepared using a similar route to as described previously (**36**). Substrates were loaded onto apo-CurM ACP using the phosphopantetheinyl transferase, Svp (**65**). The loaded carrier protein was incubated along with the sulfate donor, 3'-phosphoadenosine-5'-phosphosulfate (PAPS), as well as CurST, CurTE or both enzymes and the reactions monitored by LC-MS. Incubation of CurM ACP alone as a negative control revealed a mixture of loaded- and holo- carrier protein and gave a baseline for the downstream analysis (Figure 44). Addition of the CurTE to the reaction mixture resulted in no change (data not shown). Incubation of the loaded ACP and CurST resulted in the formation of a new peak with an 80Da mass shift, in agreement with the β -sulfonation (Figure 45). A mixed population of the holo-, loaded-, and sulfo- species as discernable, suggesting the ST catalyzed transformation did not go to completion. By HPLC, the ST-TE containing reactions demonstrate a marked shift in the retention time of the carrier product (Figure 46), indication that the enzymes are functioning in tandem, though in agreement with the investigation of the TE functionality, only the hydrolysis product is formed.

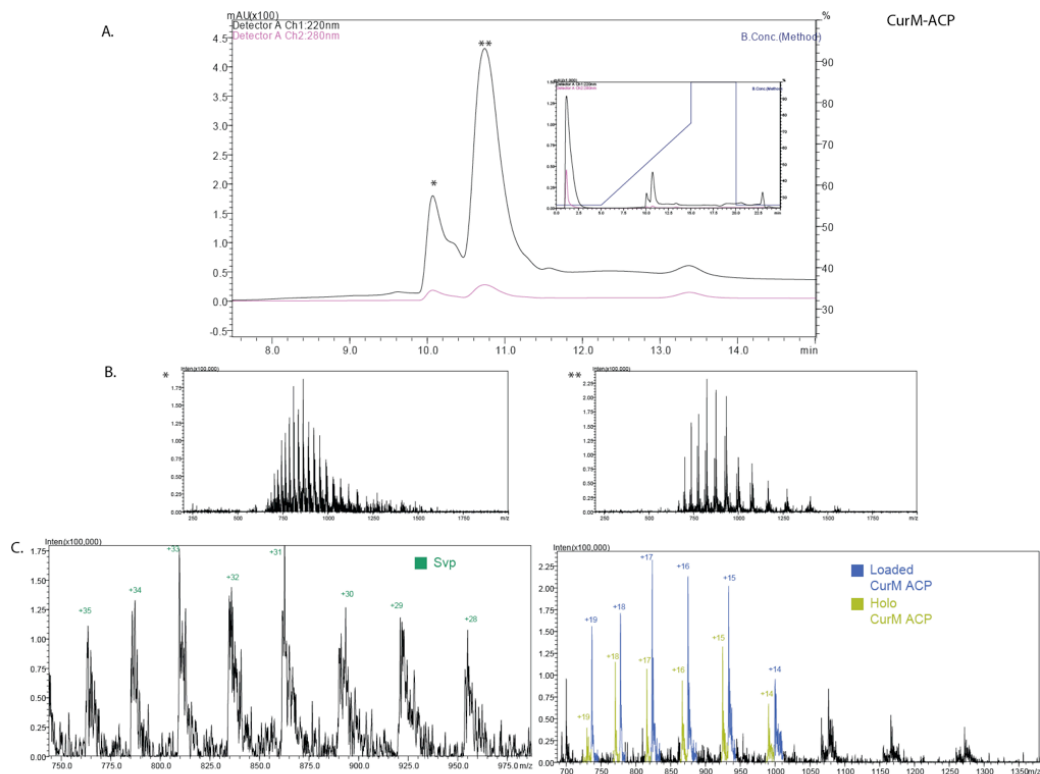


Figure 44. LC-MS analysis of the product of CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA. A. Zoomed UV chromatogram; black trace 220nm, pink trace 280nm; inset full chromatogram. B. Average mass spectrum of protein peak denoted by asterisks. C. Charge state assignment for Svp and ACP isoforms.

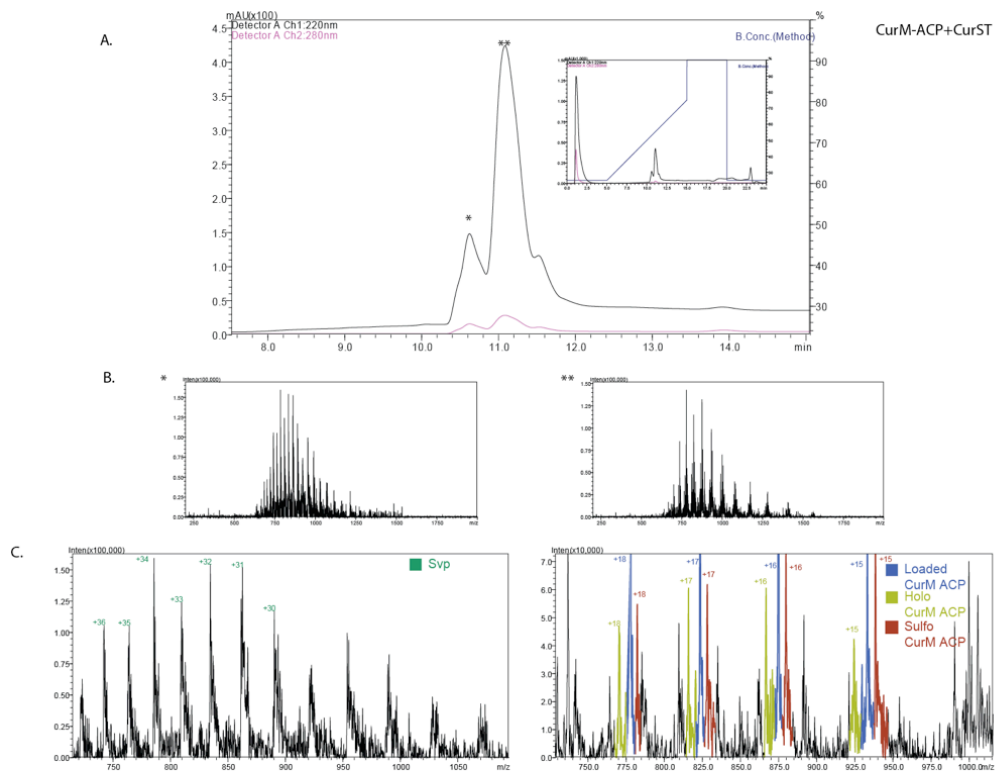


Figure 45. LC-MS analysis of the product of CurST and CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA. A. Zoomed UV chromatogram; black trace 220nm, pink trace 280nm; inset full chromatogram. B. Average mass spectrum of protein peak denoted by asterisks. C. Charge state assignment for Svp and ACP isoforms.

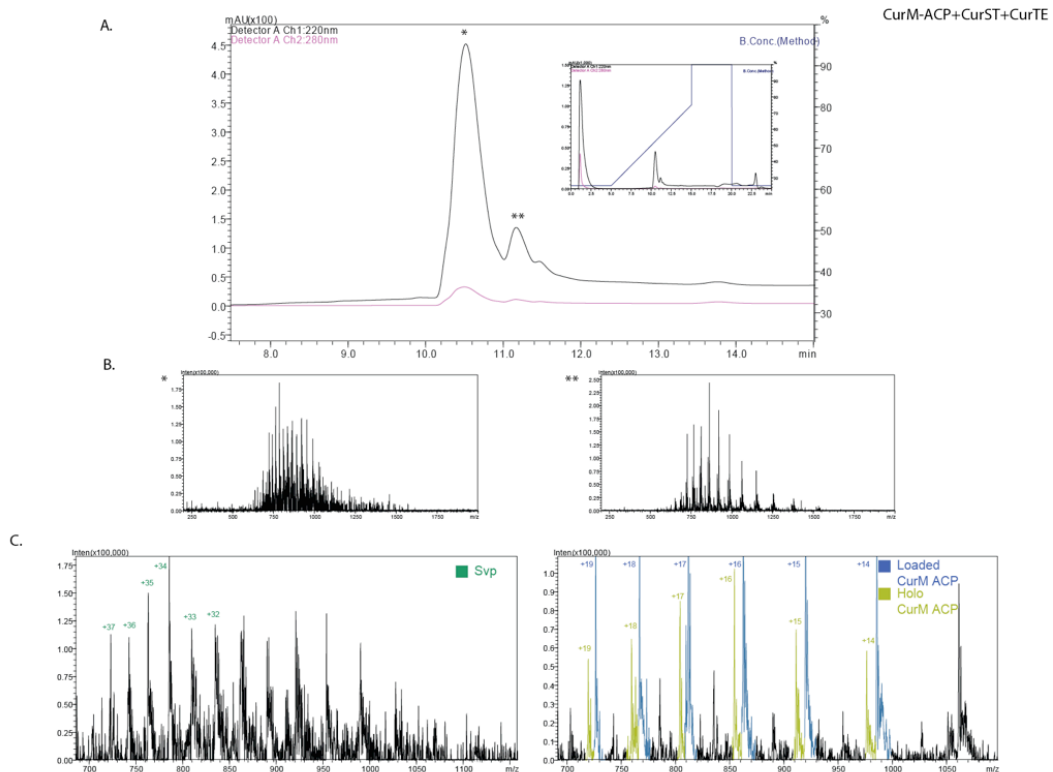


Figure 46 LC-MS analysis of the product of CurST, CurTE CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA. A. Zoomed UV chromatogram; black trace 220nm, pink trace 280nm; inset full chromatogram. B. Average mass spectrum of protein peak denoted by asterisks. C. Charge state assignment for Svp and ACP isoforms.

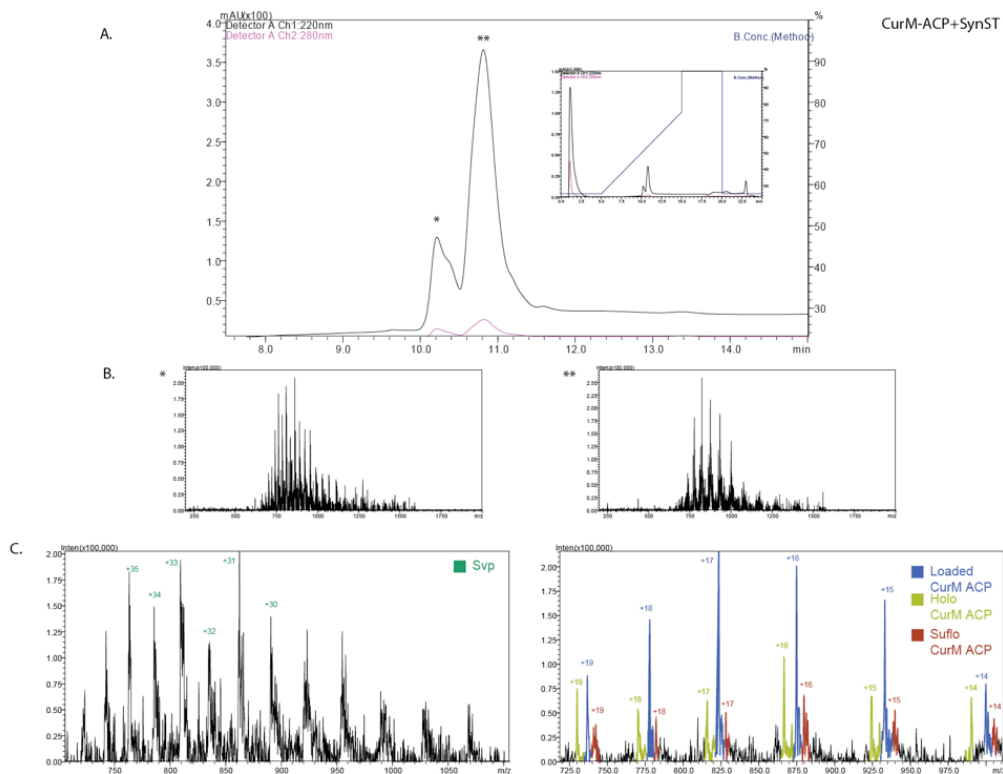


Figure 47. LC-MS analysis of the product of SynST and CurM ACP loaded with 3-hydroxy-tetradecanoyl-CoA. A. Zoomed UV chromatogram; black trace 220nm, pink trace 280nm; inset full chromatogram. B. Average mass spectrum of protein peak denoted by asterisks. C. Charge state assignment for Svp and ACP isoforms.

Having demonstrated the ability of the CurST domain to sulfonate the β -alcohol of nonnative-acyl chains, we sought to probe the capacity of the domains from the Syn and Cur system to act *in trans*. Remarkably, both CurST and SynST were capable of interacting with non-cognate ACPs (Figure 47), however the Syn sulfotransferase demonstrated approximately 20% activity as compared with Cur enzyme. Interestingly, while the SynST was capable of interacting with the Cur proteins, the enzyme was incapable of accepting the CurM-substrate harboring the 5-methoxy for sulfonation as determined by HPLC (66).

3.2.5 ST structure and loop exchange

In order to understand the observed substrate intolerances, and gain access to more fuel-like molecules through the functioning of ST-TE enzyme pair, we undertook a structural investigation of both the Cur and Syn STs (66). These enzymes represent a novel subclass of sulfonating proteins that possess a unique fold, surface substrate binding cleft, as well as a dynamic active site flap. Mechanistically, the CurST and SynST also share a common 3'-phosphoadenosine-5'-phosphosulfate binding mode that was revealed through electron density at

the active site with $ZnCl_3$, as the PAPS SO_4 surrogate. The major distinguishing element between the proteins was the active site flap, which remained unstructured in the SynST crystal though ordered for CurMST. The role of this 13 amino acid region in substrate recognitions was determined by generating chimeric forms of each enzyme, by exchanging the respective flap regions. For the CurMST_{flapSyn}, the activity towards the β -hydroxy substrates remained unchanged, while the ability to accept methoxy-bearing substrates was reduced by 50%. The SynST_{flapCurM} showed a near complete loss of activity for the β -hydroxy substrates, however, remarkably, the enzyme was now tolerant the CurM 5-(*R*)-methoxy-3-(*R*)-hydroxy-tetradecanoyl-ACP substrate. Taken together these results demonstrated the ability of the sulfotransferase to selectively accept substrate is partially conferred through the active site flap.

3.2.6 Chemical and enzymatic access of acyl-CoA substrates

Seeking to further expand our substrate pool for examining ST and TE activities, alternative methods for generating acyl-CoAs were examined. Synthetic preparation of substrates for this study, first utilized methods developed by Walsh and coworkers, wherein acyl-acids are ligated to CoA, N-acetylcysteamine or phosphopantetheine or other chemical ACP surrogates using PyBOP as the coupling agent, di-isopropylethylamine (DIPEA) as the base and an excess of thiol ranging from 2-10 molar equivalents (67). Recent studies using full-length chain elongation intermediates prepared as thiophenol thioesters, demonstrate chain transfer of substrate by processing domains and conversion by downstream PKS and NRPS enzymes; this study suggested the ability to overcome the need for CoA loading onto ACPs for downstream ST-TE reaction. Alternatively, Padamakumar *et al.* (68) have prepared malonyl-thiophenolates for a direct transthioesterification onto CoA. Though thiophenolate adducts are obtainable for 3-OH fatty acids ranging from 8-12 carbons, the amphipathic longer chain substrates makes the second step of the reaction implicitly more challenging. This is consistent with yields decreasing from 90-60% between C-12 to C14 carbons chain lengths. Moreover, 0% conversion is observed for C16-18 carbon chains.

The inability to chemically prepare longer chain acyl-CoA substrates, prompted us to investigate the enzymatic preparation of acyl-CoAs. The mycobacterial acyl-CoA ligase, FadD6, was previously shown to activate free fatty acids directly as CoA thioesters in an ATP dependent

reaction (69, 70). Several constructs of the mycobacterial FadD6 were graciously donated by RS Gokhale. As the role of this enzyme during the research was preparative, we screened each form of the enzyme for ease of overexpression, purification and highest activity. The MBP-FadD6 proved to give the highest yield and was chosen for downstream application. Fatty acids of varying length were incubated with FadD6. Reactions were quenched removed of insoluble material and analyzed by HPLC. While FadD6 demonstrated a substrate profile similar to what was previously reported (69, 70), capable of accepting fatty acid ranging in chain length and oxidation at the C-3 position, the enzyme showed significantly lower turnover for C14-C18 fatty acids. Additionally, functionalization of the longer chain acids with a β -alcohol resulted in an even lower turnover. The high cost of non-recoverable starting material from this reaction and large scale required to access minute amounts of β -hydroxy acyl CoAs and instability during purification led us to abandon this method.

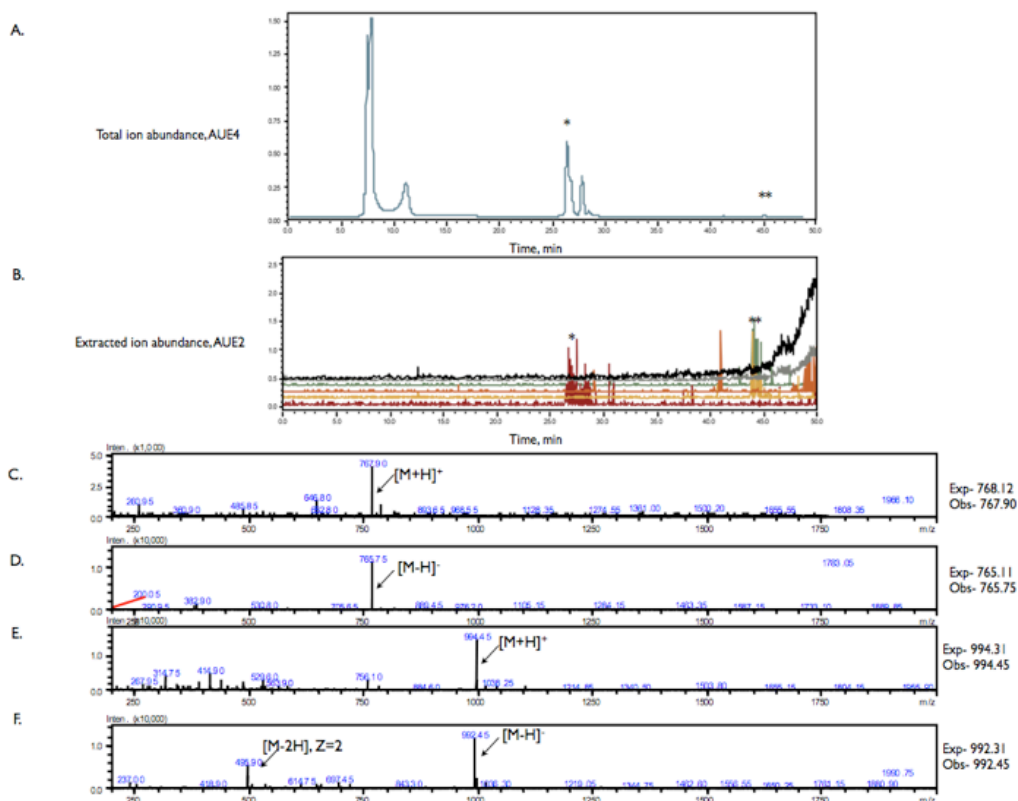


Figure 48. LC-MS analysis of FadD6 catalyzed formation of 3-OH-tetradecanoyl-CoA. A. Total ion chromatogram. Asterisk denotes unreacted CoA. B. Extracted ion chromatogram. Red trace and orange trace, selected ion monitoring for CoA in positive and negative mode. Yellow and green trace selected ion monitoring of 3-OH-tetradecanoyl-CoA in positive and negative mode. C. Mass spectral average for CoA in positive mode. D. Mass spectral average for CoA in negative mode. E. Mass spectral average for 3-OH-tetradecanoyl-CoA in positive mode. F. Mass spectral average for 3-OH-tetradecanoyl-CoA in negative mode.

3.2.6.1 Acyl activating enzymes couple free fatty acids to holo-carrier proteins

Throughout this study, CoA substrates were loaded using Svp onto apo- carrier protein in an *in vitro* biochemical surrogacy. The high cost and synthetic or enzymatic difficulties encountered while producing these activated thioesters, led us to consider utilizing enzymes in the direct acylation of holo- carrier proteins with a nonactivated (i.e. free acid) substrate, bypassing the need for substrate derivitization and the need for a loading step.

An examination of the related cyanobacterial ST-TE containing gene clusters, revealed a protein not found within CurM, annotated to be an acyl activating enzyme (AAE). BLAST analysis indicated that this protein belonged to the fatty acyl-AMP ligase (FAAL) superfamily, which utilizes ATP to catalyze the formation acyl-adenylates. Enzymatically, this class of proteins share incredibly high similarity to the related fatty acyl-CoA ligase (FACL), which like FAALs form, acyl-AMP intermediates, but follow this by a second round of catalysis transferring the substrate to the terminal thiol of CoA, or to a holo-carrier protein. This class of enzymes is found in principally in secondary metabolism, having been recently characterized in phthiocerol dimycocerosate (71) and mycolic acid (72) biosynthesis.

A global alignment of several FAAL as well as FACLs, show modest sequence homology (~20.2%), with a highly conserved nucleotide-binding motif. Interestingly, the *Synechococcus* PCC 7002 acyl activating enzyme possess the distinguishing insertional region between the N- and C- terminal subdomains (Figure 49) (73), clading the protein firmly within the FAAL subclass of Adenylate-forming-domain proteins. This finding would suggest that the protein activity would stall following acyl-adenylate formation, rather than loading onto the adjacent ACP.

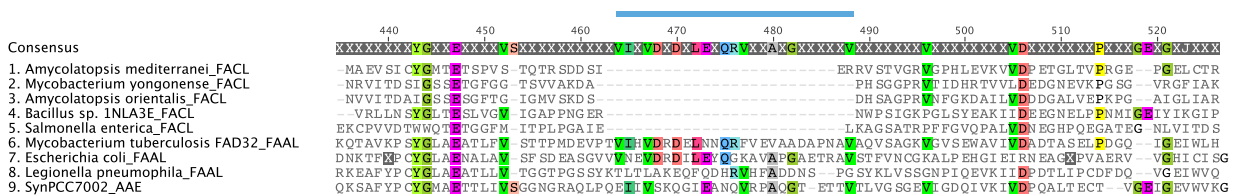


Figure 49. Alignment of fatty acyl CoA ligases and fatty acyl AMP ligases. The common interdomain insertion distinguishing FACLs from FAALs, denoted by a blue bar, is also present in the Syn acyl activating enzyme.

The AAE was cloned from genomic *Synechococcus* PCC7002 DNA as an excised domain, purified to homogeneity and the identity of the protein was confirmed by mass spectrometry. To test the activity of the enzyme, AAE was incubated with ATP, Mg^{2+} , stearic acid, and holo-CurM ACP and modifications to the ACP were monitored by MS. To our surprise, under all conditions tested, no modifications to the CurM ACP were detected.

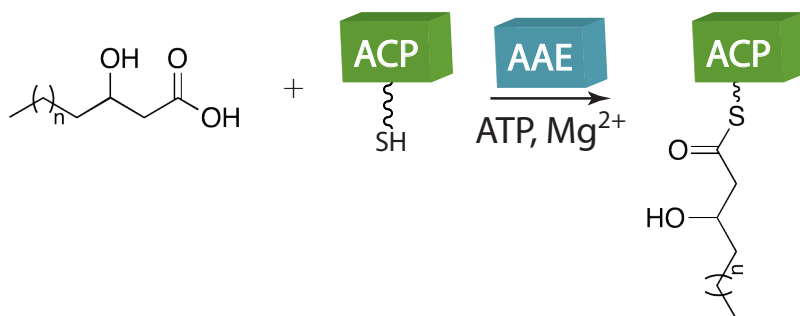


Figure 50. Reaction scheme for examining AAE activity. Conversion of holo-ACP to fatty acid loaded carrier protein was monitored by LC-MS.

Analysis of the literature regarding secondary metabolic activating enzymes indicated that a common attribute of FACs is the ability of the protein to discriminate amongst cognate ACPs (72, 74, 75). This feature enables interactions to occur at precise timing during the biosynthesis of secondary metabolites, raising the potential that the AAE would specifically load the native downstream ACP. Alignment of the CurM ACP as well as the 2 Syn ACPs demonstrates the dissimilarity of the proximal carrier protein to the distal carrier proteins and its potential involvement in the AAE specific interaction (Figure 51).

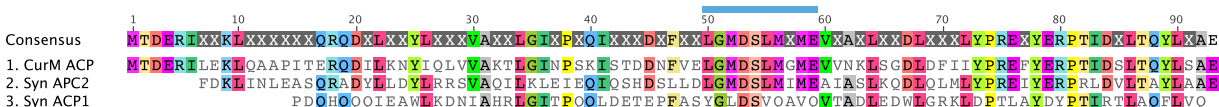


Figure 51. Alignment of ACPs from CurM and Syn olefin synthase. Blue bar indicates highest conserved region and potential ACP recognition interface.

Two constructs of the SynACPI were cloned from genomic DNA and were designed to include the minimally conserved ACP domain, DQH construct, or include a long N-terminal region that includes a portion of the linker domain between the AAE and the ACP, LDG construct.

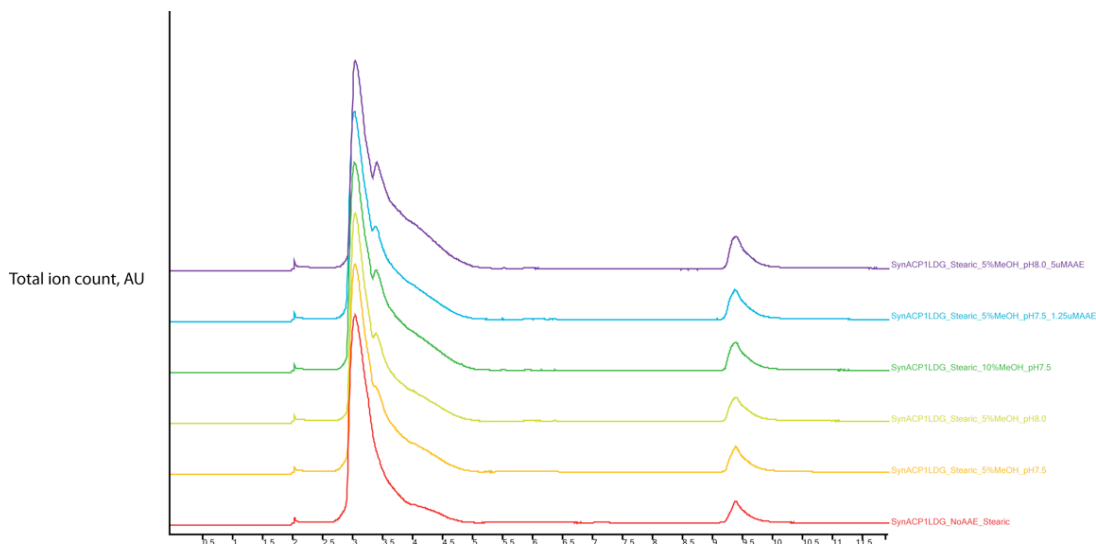


Figure 52. TIC shows formation of new peak upon incubation of Syn AAE with Syn ACP1. Asterisk denotes newly formed peak.

The ACPs were prepared in the holo-form and confirmed by MS. The AAE and SynACP1 constructs were incubated along with stearic acid and reaction cofactors as previously described, and monitored by MS. For the shorter DQH construct, no modifications to the ACP were observed, however, analysis of the total ion chromatogram (TIC) showed a new peak with slightly increased retention time, appearing upon incubation of AAE and LDG-ACP (Figure 52).

Deconvolution of the mass spectrum revealed a new peak with a mass increase of 267 Da, relative to WT, in agreement with the addition of stearic acid (Figure 53).

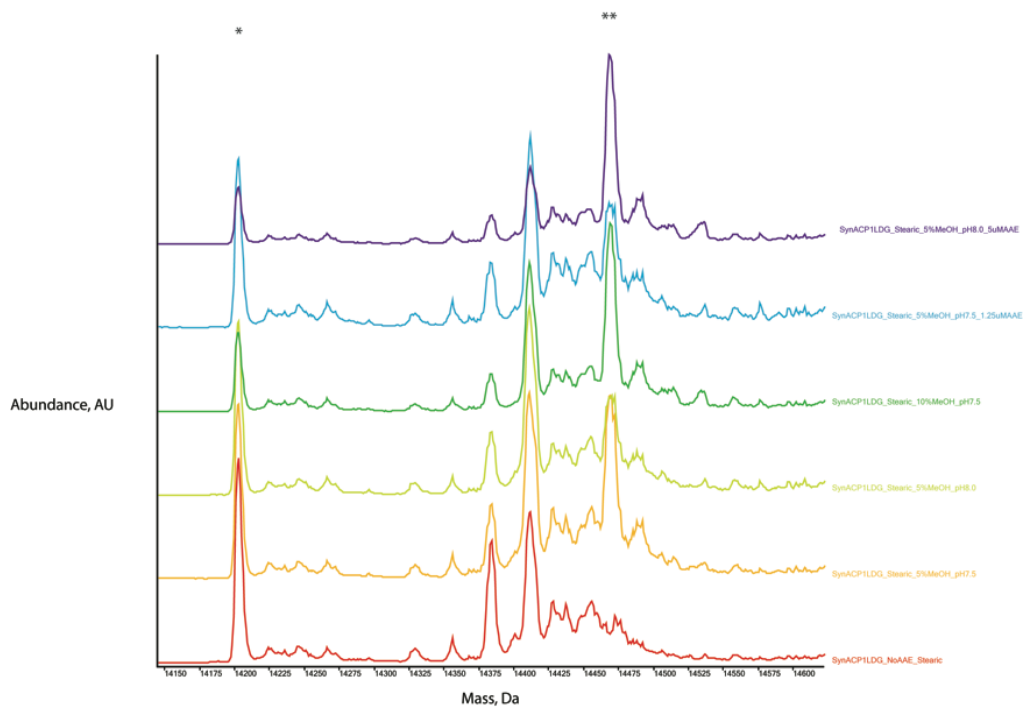


Figure 53. Deconvoluted mass abundances for AAE loaded SynLDG-ACP . Indicated with a single asterisk is the holo-carrier proteins, Calc- 14193, Obs- 14194; two asterisks denotes stearate loading, Calc- 14461, Obs- 14462.

Localization of the mass increase was determined to be at the PPant prosthetic group on the ACP by comparison of the PPant ejection ion (PPEI) products of the negative control lacking the AAE, and reactions containing the enzyme (Figure 54).

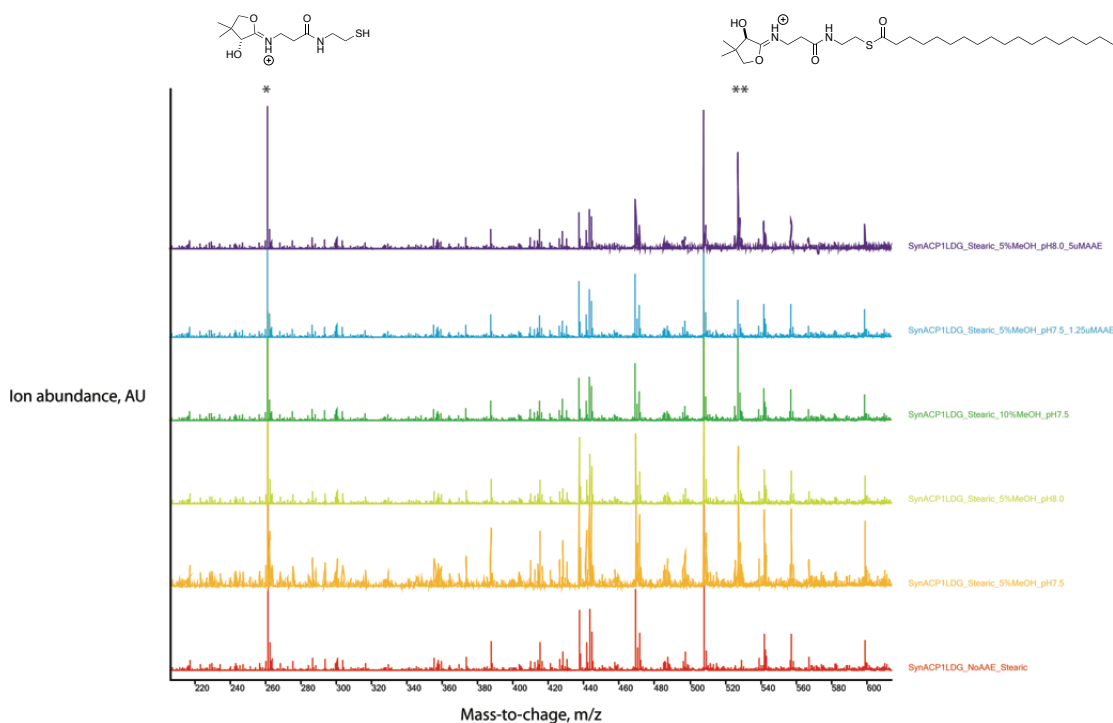


Figure 54. Zoom of average mass spectrum from 3.330-3.586 min for AAE loaded SynLDG-ACP. Indicated with a single asterisk is the holo- PPEI, Calc- 261.12, Obs- 261.2; two asterisks denotes stearate loading, Calc- 527.3, Obs- 527.3.

To maximize ACP loading, reaction conditions were scouted to increase the availability of the substrate by dissolving the free fatty acid in a variety of protein amenable cosolvents such as methanol, ethanol, and DMSO. Though the AAE is an incredibly well behaved protein in solution, it also appeared to be the least stable component of the reaction mixture with precipitation occurring under all conditions tested.

3.3 Discussion

The unique alkene forming off-loading mechanism, first characterized in the Curacin pathway represents a novel method for the biosynthetic production of terminally olefinated small molecules. Domains from the related pathway from *Synechococcus* PCC 7002, were demonstrated to have similar activity, though less tolerant to substrate functionalization than the corresponding Cur proteins. The full length *Synechococcus* Ols protein was recently overexpressed in *E. coli* (76) though no alkene production was reported. This result is in-line with the undetectable levels of hydrocarbon generated from the excised domains under *in vitro* conditions. In a typical *in vitro* assay, the ACP-loaded substrate is present at 100-300% higher

concentration than the TE domain. It is possible that the equilibrium for the second half of the TE functionality is strongly perturbed by the excess of substrate and that by limiting the amount of β -sulfo-acyl-ACP, higher titers of the alkene can be generated. Examining the activity of single turnover reactions using an ACP-ST-TE tridomain may demonstrate the 1:1 ratio of ACP to TE is critical for hydrocarbon production. If this mode is demonstrated to be successful in alkenes formation, the single turnover nature of the tridomain can be overcome by replacing the carrier protein with an ACP capable of being loaded by an acyl activating enzyme, thus generating a platform for the direct conversion of free fatty acids to alkene hydrocarbons.

As with other biofuel alternatives, critical barriers to the implementation of pioneering technologies are overcoming potential socio-political challenges. The discourse surrounding bioenergy alternatives continues to largely oppose the supplantation of fossil fuels. This resistance comes from the gas and petroleum industries in addition to domestic and foreign policy makers. Of principal concern for the long-term adoption of biofuels are the translational costs of utilizing biomass for fermentation, the need to curb greenhouse gas production and the ability to keep pace with increasing international energy demands. Global competition for arable land in either food production or as biofuel feedstocks has led to a rise in land costs along with the price of major agricultural commodities such as corn and soybeans. Within our developing biofuels system, this imbalance is overcome by virtue of the native biology of aquatic photoautotrophs from which the olefin synthase genes of interest originate. Blue-green algae are readily cultivable on nonarable land and in non-potable water, thus ameliorating concerns over the appropriation of farmable land and limited water resources. Additionally, the inherently higher photosynthetic-efficiency of cyanobacteria contrasted with terrestrial plants, leads to an increased capture of CO₂ from the atmosphere, which has correlated directly with increased growth rates and higher hydrocarbon titer per gram of biomass. A comparison study of the total usable energy output from microalgae has demonstrated a 2,000 – 16,700% greater oil production efficiency than traditionally derived biomass on a per hectare basis, firmly establishing cyanobacteria as valuable biofuel alternative.

Unlike current alternatives to fossil fuels, major infrastructure and economic challenges are also mitigated using an algae-based biofuel platform. The energy balance for ethanol derived from

corn is known to be net negative, primarily due to expenses associated with industrial production using corn feedstocks and the large energy input required for the separation and isolation of ethanol. Furthermore, the lower energy density and high hygroscopicity of ethanol renders it unusable in current engine and distribution networks. In contrast, the nature of the liquid fuel generated from our enzyme system make them directly compatible with the current fossil-fuel infrastructure as a drop-in alternative, thus obviating the need to construct a biofuels-compatible framework.

3.4 Materials and methods

Bacterial strains, media and culture conditions.

Escherichia coli DH5 α subcloning efficiency cells (Invitrogen) were used as axenic stocks for the preparation of electrocompetent cells for DNA propagation and maintenance. *Escherichia coli* Rosetta BL21(DE3) (Novagen) harboring the pRARE plasmid (spectinomycin resistance exchanged for chloramphenicol by Dr. W. Clay Brown, University of Michigan) for general protein overexpression; holo- carrier proteins were obtained using the BAP1 cell line (Pfeifer).

All transformations were achieved via electroporation (BioRad, Gene Pulser Xcell), with 2.5-500 ng DNA from ligation reaction mixture or as intact plasmid, recovered with SOB, allowed to outgrow for 60 minutes prior to plating on lysogeny broth-Miller-agar plates (LB-Miller) supplemented with the appropriate antibiotic (ampicillin- 50 μ g/mL, kanamycin- 50 μ g/mL, spectinomycin- 25 μ g/mL).

Protein overexpression, media and culture conditions

Fresh transformants harboring the plasmid of interest were used to inoculate a culture of lysogeny broth-Miller (LB-Miller), corresponding with plasmid resistance marker (ampicillin- 50 μ g/mL, kanamycin- 50 μ g/mL, spectinomycin- 25 μ g/mL). Overnight (~12-18 hours) cultures were used to subinoculate (1:100 (v/v)) large-scale (500-1000 mL) cultures supplemented with appropriate antibiotics for protein overexpression. Apo- carrier proteins were cultured in LB-miller supplemented with 1% (v/v) Glycerol, 25mM Tris (pH= 7.5), and freshly prepared 100 μ M Fe(II)SO₄, while the ST, TE AAE and SVP were cultured in LB without modification.

For all proteins of interests, large-scales cultures were grown at 37° while shaking at 160 RPM until reaching an OD₆₀₀ 0.6-0.8. Flasks were allowed to cool at room temperature for 20-30 minutes. Protein overexpression was induced with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.3-0.4 mM. Flasks were returned to a 20° shaker and continued to grow. Apo- carrier proteins were harvested following 5 hours induction, while ST, TE AAE and SVP were allowed to grow for 12-16 hours prior to harvesting. Cells were harvested by centrifugation at (5000xg, 10 minutes, 4°), transferred to 50 mL Falcon tubes and cell pellets were flash frozen in liquid nitrogen and stored at -80° until purification.

Protein purification

All steps during the protein purifications were performed at 4°. Cell pellets were resuspended in 2x cell volume (minimum 25mL) 5-10% Buffer B, and incubated with freshly prepared lysozyme (0.2 mg/mL final, prepared in MilliQ H₂O), freshly prepared PMSF (1 mM final, prepared in DMSO) and 0.5 *cOmplete* protease inhibitor cocktail (Roche) tablet for 30 minutes while rocking. Cells were lysed by use of a recently tuned ultrasonic dismembrator for 10 cycles of 1 minute, with pulses of 5 seconds on, followed by 5 seconds off, and intensities ranging from 7-8. All cycles were carried with chilled beakers in packed ice bucket. Cellular debris was removed by centrifugation at 40000-50000xg for 30 minutes and the supernatant was transferred to a fresh tube; residual insoluble material was removed by filtration (syringe driven 0.45 μm filter) or a second round of centrifugation.

Immobilized metal affinity chromatography

Clarified lysate was applied to 2 mL nitrotriacetic acid-agarose resin (Qiagen), charged with Ni²⁺, and equilibrated with 10-20% Buffer B. The resin was washed extensively with 25 column volumes of 15-30% buffer B to remove any nonspecifically bound proteins. The target protein was eluted with a stepwise gradient from 20-100% Buffer B. Peak fractions as determined by gradient SDS-PAGE ((4-12%) (Novex NuPage) run in MES buffer at 180 mV for 50 minutes) were pooled and concentrated using Amicon ultra centrifugal concentrators (Millipore, 0.5 kDa, 3 kDa, 10 kDa), equilibrated with Buffer F until a volume of 2.5 mL was reached. Proteins were buffer exchanged with a PD10 washed and equilibrated with Buffer F and eluted with 3.5 mL of

Buffer F. Proteins were further concentrated to less than 100-500 μL , aliquoted, flash frozen in liquid nitrogen and stored at -80° . Protein concentrations were determined using the bicinchoninic acid assay (BCA Pierce) according to the manufacturers guidelines, using a BSA standard curve and read in 96 well format.

Preparation of substrate-loaded ACP

Apo- carrier proteins, as determined by mass spectrometry, were loaded with either (*R*) 3-hydroxy-tetradecanoyl Coenzyme A, or (*R*) 5-methoxy-3-hydroxy-tetradecanoyl Coenzyme A. Briefly, in a 1000 μL reaction (50 mM Tris (pH= 8.0) 50 mM NaCl, 1 mM MgCl_2 , 10% (v/v) glycerol, 1 mM TCEP)), 250-500 μM ACP was incubated along with 5 μM recombinant *Streptomyces versicolor* phosphopantetheinyl transferase, SVP. The reaction was initiated with the addition of 10 molar equivalents of substrate-CoA prepared as a 5 mM stock in MilliQ H_2O . Reactions were incubated at room temperature for 2 hours followed by dilution to 1.5mL with Buffer F. The diluted reaction was loaded onto a HiTrap Desalting column (GE Healthcare) equilibrated with Buffer F and eluted with 2 mL Buffer F, to remove unreacted CoA, TCEP and salts. Proteins were concentrated, aliquoted, flash frozen in liquid nitrogen and stored at -80° . The concentration was measured and extent of substrate loading was determined by mass spectrometry as above.

Mass spectrometric analysis of proteins by LC-ESI-MS

Excised domain, as well as coupled, Cur and Syn ST-TE reactions were carried out as described in Gehret *et al.* (**60**, **66**). The 300 μL reaction was partitioned into 100 μL aliquots for analysis by HPLC, LC-MS, GC-MS (see below). AAE reactions were carried out in 50 μL volumes and processed similarly for exclusively LC-MS analysis.

The identities of the proteins in this study as well as the extent of phosphopantetheinylation of ACPs except for sulfonation experiments (see below) were determined by electrospray-time-of-flight (TOF) mass spectrometry. Protein samples were removed from the concentrated fraction prior to buffer exchange (minimum 2 mg/ml), diluted 1:10, acidified with 10% (v/v) formic acid (FA), centrifuged at 20800xg for 10 minutes and transferred to a fresh tube. Of this sample, 20 μL was injected onto a C4 column (Phenomenex Aeris™ 3.6 μm WIDEPOR C4 300 Å, LC

Column 50 x 2.1 mm) equilibrated with 30% CH₃CN+0.1% FA, connected online with an Agilent 6250 quadrupole – Time-of-Flight mass spectrometer. Column elution was diverted to waste for sample desalting for 1 minutes prior to a gradient of 30-95% CH₃CN+0.1%FA over 8 minutes, followed by a 1 minute wash at 95% CH₃CN+0.1%FA and reequilibration at 30% CH₃CN+0.1%FA for 2 minutes. Alternatively a PLRP-S column (Varian 3 μm, LC Column 50 x 4.6 mm) equilibrated with 10% CH₃CN+0.1% FA was connected online with a Shimadzu LC-MS 2010 EV (single quadrupole, ESI configuration operated in the positive, profile mode). Column elution was diverted to waste for sample desalting for 5 minutes prior to a gradient of 10-95% CH₃CN+0.1% FA over 20 minutes, followed by a 5 minute wash at 95% CH₃CN+0.1%FA and reequilibration at 10% CH₃CN+0.1%FA for 10 minutes.

Spectra acquired using TOF were processed using the Bioconfirm Intact Protein module on the Agilent MassHunter Qualitative Analysis software. Protein identities were confirmed by obtaining an averaged mass spectrum over maximum peak height as determined by Total Ion Chromatogram (TIC) with a 30 second background spectrum subtracted. Spectral deconvolution was performed over a range of 200-2000 m/z, using a minimum of 7 consecutive charge states and fit score of 8, with a limited mass range of ±10% of the calculated mass of the protein. Experiments using the single quadrupole mass spectrometer were deconvoluted by hand using the calculated average mass mass-to-charge ratio for individual proteins as a reference for peak assignment.

GC-EI-MS analysis of hydrocarbon production

TE catalyzed olefinative offloading was determined using a protocol adapted from conversations with Dr. Liangcai Gu. The 100 μL reaction aliquots (see above) were diluted to 500 μL with MilliQ H₂O and extracted thrice with an equal of ethyl acetate (EtOAC). The samples were dried under nitrogen gas and subsequently reconstituted in 500 μL EtOAC. GC-MS assays were conducted using a Shimadzu QP-2010 GC-MS single quadrupole mass spectrometer. GC methods were developed using acyl-chain standards, systematically worked-up as with enzymatic reactions. Maximal allowable injections of 8 μL (splitless) were made onto a DB-5 column (Agilent, 30 m x 0.25 mm) operating with helium-carrier gas. The injector and detector temperatures were set at 200°C. After an initial setting at 30°C for 2 minutes, the oven

temperature was raised to 250°C at 18.3°C /min and held for 2 minutes. Spectra covering a mass range of 50-500 m/z were collected and processed with the Labsolutions Postrun software.

Thio-click derivatization of enzymatically produced olefins

Alkene functionalization proceeded using experimental methods adapted protocols outlined in Hoyle *et al.* (64). Briefly, in a 1000 µL reaction (50 mM Tris (pH= 8.0), 300 mM NaCl, 10% (v/v) MeOH) 60 µM 1-tridecene was mixed along with 600 µM N-acetylcysteamine and 0.6 µM benzophenone (prepared in MeOH). The mixture was incubated at room temperature for 30 minutes in a foil-lined box under exposure to a handheld UV-lamp. The reaction was processed by centrifugation at 20800xg for 10 minutes to remove insoluble material and loaded onto a 1 mL disposable reverse phase octadecyl extraction column (Bakerbond, C18) equilibrated with 5% (v/v) MeOH+ 0.1% FA. The column was washed with 10 column volumes of 5% (v/v) MeOH+ 0.1% FA and subsequently eluted with a stepwise gradient of 10%-100% MeOH+ 0.1% FA in 1 mL fractions. 20 µL samples of each fraction were assayed for thio-click-product loading sample onto an XBridge C18 column (Waters, 100 mm x 4.6 mm) equilibrated with 30% MeOH connected online with a Shimadzu LC-MS 2010 EV (ESI configuration). Column elution was diverted to waste for sample desalting for 1 minutes prior to a gradient of 30-95% MeOH over 10 minutes, followed by a 10 minute wash at 95% MeOH and reequilibration at 30% MeOH for 10 minutes. Profile spectra were acquired in the positive mode over a range of 50-500 m/z using and processed in Labsolutions Postrun software.

FadD26 catalyzed acyl-CoA formation

CoA ligation reactions were carried out as described in Arora *et al.* (70) without modification. The reactions were quenched with the addition of 10% (v/v) ice-cold trichloroacetic acid and centrifuged at 20800xg for 10 minutes to remove precipitated proteins. 10 µL of the soluble fraction was loaded onto an XBridge C18 column (Waters, 100 mm x 4.6 mm) equilibrated with 10% CH₃CN (in 10 mM CH₃COONH₄ (pH= 5.4)) connected online with a Shimadzu LC-MS 2010 EV (ESI configuration). Column elution was diverted to waste for sample desalting for 5 minutes prior to a gradient of 10-95% CH₃CN over 20 minutes, followed by a 5 minute wash at 95% CH₃CN and reequilibration at 10% CH₃CN for 10 minutes. Spectra were acquired in both

the positive and negative mode over a range of 100-2000 m/z and processed in Labsolutions Postrun software.

AAE loading of holo- carrier proteins

Acyl activating enzyme loading was examined via a protocol similar to the characterization of ligases by Zhang *et al.*(73). Briefly in a 50 μ L reaction (50 mM Tris (pH= 8.0) 50 mM NaCl, 1 mM MgCl₂, 1% (v/v) glycerol, 10 mM ATP, 10% (v/v) MeOH), 50 μ M ACP was incubated along with 500 μ M stearic acid (prepared in MeOH) and 0.5-5 μ M AAE. The reactions were monitored for turbidity associated with protein precipitation during incubation at 24°C, and quenched with formic acid (10% (v/v) final concentration) at times associated with a loss turbidity increase. Quenched reactions were analyzed by LC-MS as described above.

Acknowledgements: This work was performed collaboratively between the Sherman and Smith labs. Dr. Liangcai Gu generated WT CurM ACP, ST and TE constructs, Dr. Jennifer Gehret-McCarthy (JGM) prepared mutant Cur and Syn STs and TEs and Eli Benchell Eisman (EBE) and Chris Kubitskey (CK) generated Syn AAE and ACPs constructs. Protein preparations including analytical gel filtration and substrate loading were conducted by JGM, EBE and CK. JGM and EBE designed all experiments. ST and TE experiments were prepared by JGM, FAD experiments were conducted by EBE with assistance from Jake Bermudez, AAE experiments were done by EBE and CK. Protein and small molecule LC-MS assays were conducted by EBE with assistance from CK, while protein HPLC experiments were carried out by JGM with support from EBE. EBE conducted all LC-MS and GC-MS experiments for acyl-CoA, thio-click and alkene assays. JGM and Prof. Janet L. Smith solved the crystal structures of Cur and Syn proteins.

Chapter 4

Origins of the linear lipopeptide carmabin A

4.1 Introduction

Marine cyanobacteria have been well noted for the incredible structural diversity and broad biological activities of small molecules produced within the cyanobacterial secondary metabolome (21, 77, 78). A common scaffold of these metabolites includes an acyl chain linked to a polypeptide, suggesting that the majority of these compounds are of mixed polyketide synthase (PKS) and nonribosomal peptide synthetases (NRPS) origin. Structural elaboration is achieved amongst this class of natural products by varying the chain length and level of unsaturation of the fatty acyl tail, as well as through the incorporation of the nonproteinogenic amino acids and β -hydroxy acid monomers, leading to depsipeptide bonds within the mature compound (11). Further diversity is obtained during the termination process wherein the small molecule can be either cyclized or offloaded as a linear small molecule. Interestingly, unlike canonical cyclization strategies for macrolactone biosynthesis, such as pikromycin or erythromycin, the fatty acid chain remains unincorporated into the ring system.

Within the family of linear cyanobacterial metabolites, carmabin A, structurally characterized initially in 1998 (28) and recently coisolated with the structurally homologous dragomabins and dragaonamides (79), represent a class of lipopeptides with antimalarial activity (Figure 55).

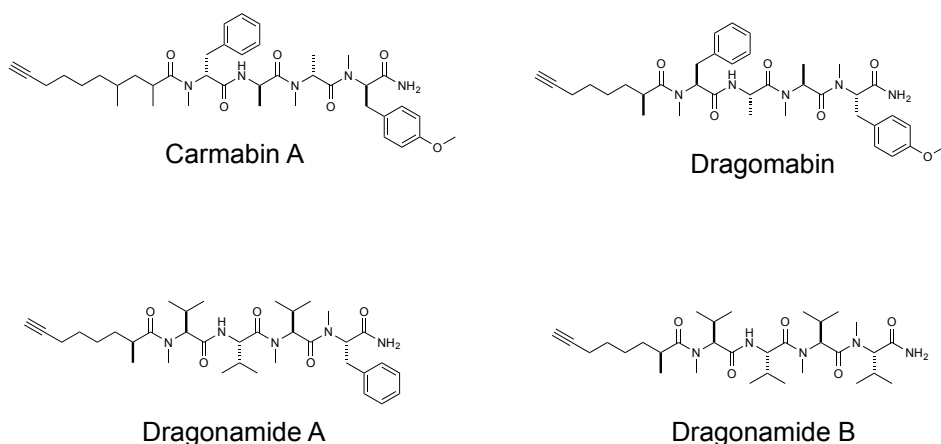


Figure 55. Cyanobacterial natural products containing an alkyne and amide.

The core scaffolds share an alkyne, attached via a C8-C10 acyl linker, to a tetrapeptide with a highly N-methylated backbone terminating in a previously unobserved amide. In addition to this small family of related molecules, the previously identified jamaicamides(27) also possess an alkyne. Given the presence of this functionality, it was thought that this chemical group would be critical to the bioactivity of the small molecule. This hypothesis allowed McPhail *et al.* (79) to glean some relationship between the structure of the metabolite and antimalarial activity. Each of the compounds were assayed for antiplasmodial activity against a known chloroquinone-resistant strain of malaria. While all four lipopeptides possessed the acetylene group, only the aromatic containing carmabin A, dragomabin and dragonamide A were efficacious with IC₅₀ values of 1.4, 21.0 and 10.7 μ M, respectively. Interestingly the activity of jamaicamide B, which has no aromatic rings, was similar to the other metabolites, 18.4 μ M, while no activity was observed for terminal bromo-acetylene containing jamaicamide A. Together this suggested that critical interactions between the natural product and target are being enabled via the alkyne. Moreover, the aromatic residues and pyrrolinone functionality of jamaicamide A, play an additional role in engendering antimalarial activity.

In addition to the terminal unsaturation, carmabin A possesses an unusual terminal amide. Reports of natural products terminating in an amide are uncommon, though those identified are typically the result of a ribosomally synthesized peptides that undergo downstream modification such as nosiheptide (80-82). Recently however, a characterization of myxathiazol

biosynthesis(83, 84) revealed an NRPS pathway in which the terminal NRPS module is dedicated to the activation of an amino acid, peptide bond formation and unusual oxidative offloading generating the observed amide. Motivated by the unique structure of antimalarial lipopeptides and we thus sought to elucidate the biosynthetic origins of both the pharmacologically significant alkyne as well as the terminal biosynthetically distinct amide.

4.2 Carmabin pathway

The carmabin sequence was identified redundantly from a screen of a cosmid library prepared from genomic *Moorea producens* 3 (unpublished work) by iterative rounds of primer-walking PCR, Sanger sequencing, and shotgun sequencing in parallel though 454 sequencing of multiple displacement amplified DNA (85). A draft carmabin cluster was mined from 272 assembled contigs and found to span the negative strand of 3 contigs. The 5'-terminus of the Car sequence was defined by the 30987nt contig AEPQ01000195, while the 3' end was defined by AEPQ01000272, which initiates at 18580bp and expands far beyond the 46710 Car terminus. The sequences of the two contigs were bridged by AEPQ01000235 the shortest assembled construct, 7282 nt, expanding between 10817- 21165. PCR revealed the complete pathway was contained over two cosmids, pCarCOS_L and pCarCOS_R, which were used for all downstream molecular biology processes.

The boundaries of the *car* gene cluster were defined by mining genes 5kbp up- and downstream of the last discernable PKS- or NRPS-like protein by screening open reading frames of 100bp for proteins putatively involved in secondary metabolism. A 4.4kbp region is 5' of the first module, which has three open reading frames (ORF) transcribed in the reverse direction ORF1 and ORF3, and one in the forward direction ORF2. ORF1 has high sequence homology with inositol monophosphatase from a related species of cyanobacterium *Microcoleus* sp. PCC 7113, ORFs 2 and 3 have unknown function but both share high amino acid identity with putative cyanobacterial proteins from *Microcoleus* sp. PCC 7113 and *Fischerella muscicola*, respectively. Directly upstream of the first module, is a 1.2kbp noncoding region that defines the 5'-cutoff. The total GC content of the Car cluster was calculated to be 41.2%, which is similar to the curacin 42.9%, (12), jamaicamide 42.1% (27), and barbamide, 45.4% (24) natural product pathways from *Moorea producens*.

The *car* cluster consists of 10 modules, denoted *carA-J*, in 10 discrete ORFs expanding 42069bp, transcribed in a uniform forward direction (Figure 56, Table 5). At the absolute 5'-end, three discrete ORFs *carA*, *B*, *C*, encode monodomain proteins that show high homology with the terminal Jam proteins. *CarA*, 96% sequence identity with *jamA*, encodes an acyl-CoA ligase, falling into the class I-adenylate forming domain superfamily, which generates AMP-activated species from free fatty acids. *CarB*, encodes a protein with putative desaturase activity, belonging to the Δ^9 -fatty acid desaturase superfamily of proteins and sharing incredibly high sequence identity, 95%, with *jamB*. The predicted protein product of *carC* is the relatively small, stand-alone acyl carrier protein. Both *carC* and *jamC* are 100 amino acids, share the invariant active site serine and differ at only 5 residues. Collectively, these early proteins constitute a common starting-tridomain system, predicted in many cyanobacterial alkyne containing natural product pathways (86).

A small intergenic break of 143bp precedes the continuation of the *car* cluster with two highly homologous PKS modules, *CarD* and *CarE* of approximately the same length, 2553 and 2576 amino acids respectively. In addition to the core ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) domains common, both *CarD* and *CarE* include the full suite of reducing enzymes, ketoreductase (KR), dehydratase (DH) enoylreductase (ER) as well as a C-methyltransferase (CMT). Both PKS proteins share very homology 69% and 70%, as well as domain architecture with *JamJ*. Major differences in sequence identity between the two *Car* proteins are relegated to the docking interfaces at the N- and C-termini. *CarF*, *G*, *H* and *I* are NRPS monomodules that range in size from 3408- 6912bp depending on domain inclusion. The minimal module *CarG*, consists of condensation (C), adenylation (A), and peptidyl carrier protein (PCP), while *CarF* and *CarH* expand domain composition to include a SAM-dependent N-methyltransferases. The largest NRPS module, *carI*, is the most genetically complex. As with the other modules, the core is comprised of a C, A, and PCP domain, *CarI* is unique in that includes two methyl transferases, catalyzing both N- and O-methylation. The most intriguing element of *CarI*, is the domain C-terminal of the PCP. The ~420 amino acid region has high sequence homology with proteins from a variety of cyanobacteria including *Cyanothece* and *Oscillatoria*, however, there is no apparent conserved function for this domain. BLAST analysis

of the enzyme indicated the protein to have reductase, dehydrogenase, or aminotransferase activity .

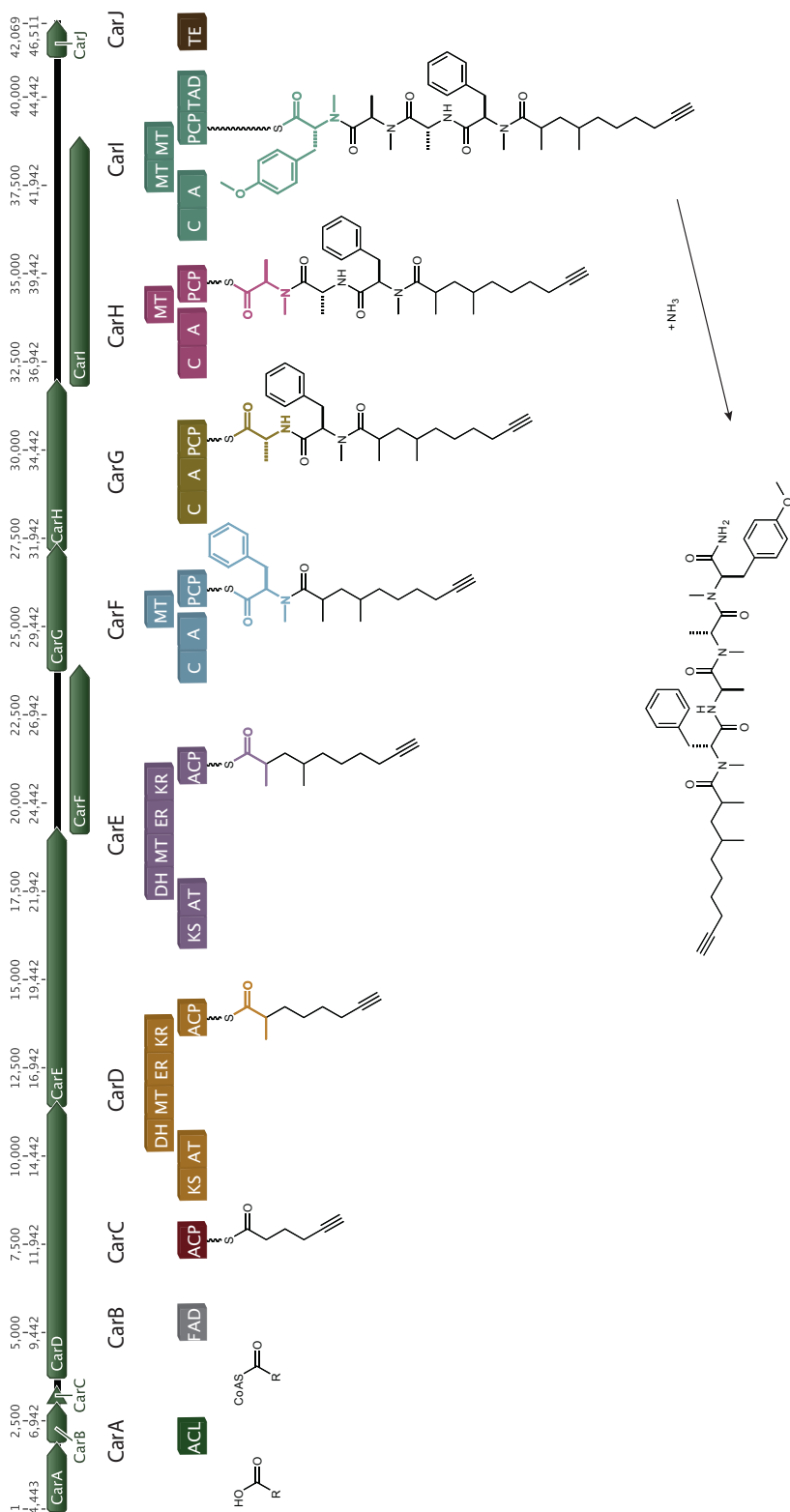


Figure 56. Proposed carmablin A biosynthetic pathway.

Table 5. Deduced functions of the Open Reading Frames in the *car* gene cluster. Domain predictions characterized by protein BLAST assignment.

Module	Domains	Start	End	Sequence similarity, Protein Origin	Identity	Similarity	Accession
Orf1	Inositol monophosphatase	18603	17737	inositol monophosphatase, <i>Microcoleus sp. PCC 7113</i>	71%	84%	YP_007120181
Orf2	UNKNOWN	18801	19229	hypothetical protein Mic7113_1167, <i>Microcoleus sp. PCC 7113</i>	63%	83%	YP_007120464
Orf3	UNKNOWN-Cupin fold superfamily	19929	19639	hypothetical protein, <i>Fischerella muscicola</i>	74%	85%	WP_016868553
CarA	Acyl-CoA ligase	22029	23810	JamA, <i>Lyngbya majuscula</i>	96%	98%	AAS98774
CarB	Fatty acid desaturase	23993	24598	JamB, <i>Lyngbya majuscula</i>	95%	98%	AAS98775
CarC	ACP	25095	25397	JamC, <i>Lyngbya majuscula</i>	95%	98%	AAS98798
CarD	KS AT DH MT KR1 ER KR2 ACP	25828	33519	JamJ, <i>Lyngbya majuscula</i>	69%	81%	AAS98781
CarE	KS AT DH MT KR1 ER KR2 ACP	33531	41259	JamJ, <i>Lyngbya majuscula</i>	70%	81%	AAS98781
CarF	C A1 MT A2 PCP	41256	45893	BarG, <i>Moorea producens 3L</i>	69%	81%	AEE88297
CarG	C A PCP	45890	49297	adenylation domain-containing protein, <i>Nostoc punctiforme PCC 73102</i>	59%	75%	YP_001869919
CarH	C A1 MT A2 PCP	49300	53940	BarG, <i>Moorea producens 3L</i>	53%	67%	WP_002733092
CarI	C A1 MT MT A2 PCP TAD	53937	60848	hypothetical protein, <i>Scytonema hofmanni</i>	59%	73%	WP_017742648
CarJ	TE(Type II)	63259	64140	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase, <i>Microcoleus vaginatus</i>	76%	81%	WP_006633653

Amino acids activated by A-domains within each NRPS module were determined by comparison of the protein sequence with domain specificities code (87). CarF likely incorporates a phenylalanine, CarG and CarH are both predicted to add alanine, while CarI has the highest agreement with tyrosine.

Table 6. Carmabin adenylation domain specificities profiles. NRPS sequences were compared with amino acid recognition as determined by Stachelhaus. Numbers as denoted by Stachelhaus. Colors denote proximity of expected and observed amino acids. Black- identical amino acid at the predicted position, blue-±1 amino acids, red- ±4 amino acids, green- >7.

Module	235	236	239	278	299	301	322	330	331	517
CarF (555-834)										
Exp- Phenylalanine	D	A	W	T	I	A	A	V	C	K
Obs	D	A	W	T	I	V	V	A	C	K
CarG (217-564)										
Exp- Alanine	D	L	L	F	G	I	A	V	L	K
Obs	D	G	L	W	G	S	G	D	S	L
CarH (274-552)										
Exp- Alanine	D	L	L	F	G	I	A	V	L	K
Obs	D	L	L	G	G	G	I	V	M	A
CarI (736-1018)										
Exp-Tyrosine	D	A	S	T	V	A	A	V	C	K
Obs	D	A	S	T	I	V	N	V	C	R
CarI TAD										
OMe-Tyrosine	D	A	L	G	V	V	G	G	L	V

4.3 Proposed carmabin A biosynthesis

The high degree of colinearity between the core structure of the mature metabolite and biosynthetic operon led to a proposal for carmabin construction, excluding termination (Figure 56). Biosynthesis of carmabin is likely initiated by oxidation of hexanoic acid to 5-hexynoic acid, catalyzed by CarB. The alkynic acid is then activated by the acyl-CoA ligase, CarA, as the acyl-AMP and subsequently loaded on the downstream ACP, CarC. Two rounds of canonical PKS elongation and C-methylation are carried out by monomodules CarD and CarE yielding the 2,4-dimethyldec-9-ynoyl-ACP. CarF marks the transition between PKS and NRPS biosynthesis, which serves to accept the acyl-substrate from the upstream module and catalyzes the additions of phenylalanine as well as methylation of backbone amide. Alanine is added sequentially by

both modules CarG and CarH, however the N-MT domain of CarH modifies the amino acid by methylating the β -nitrogen. The final module CarI incorporates tyrosine into the molecule and catalyzes methylation of amide as well the tyrosine para-hydroxy. Transamination and release of the natural product is unclear, but potentially involves the terminal CarI terminal domain.

4.4 Probing alkyne formation

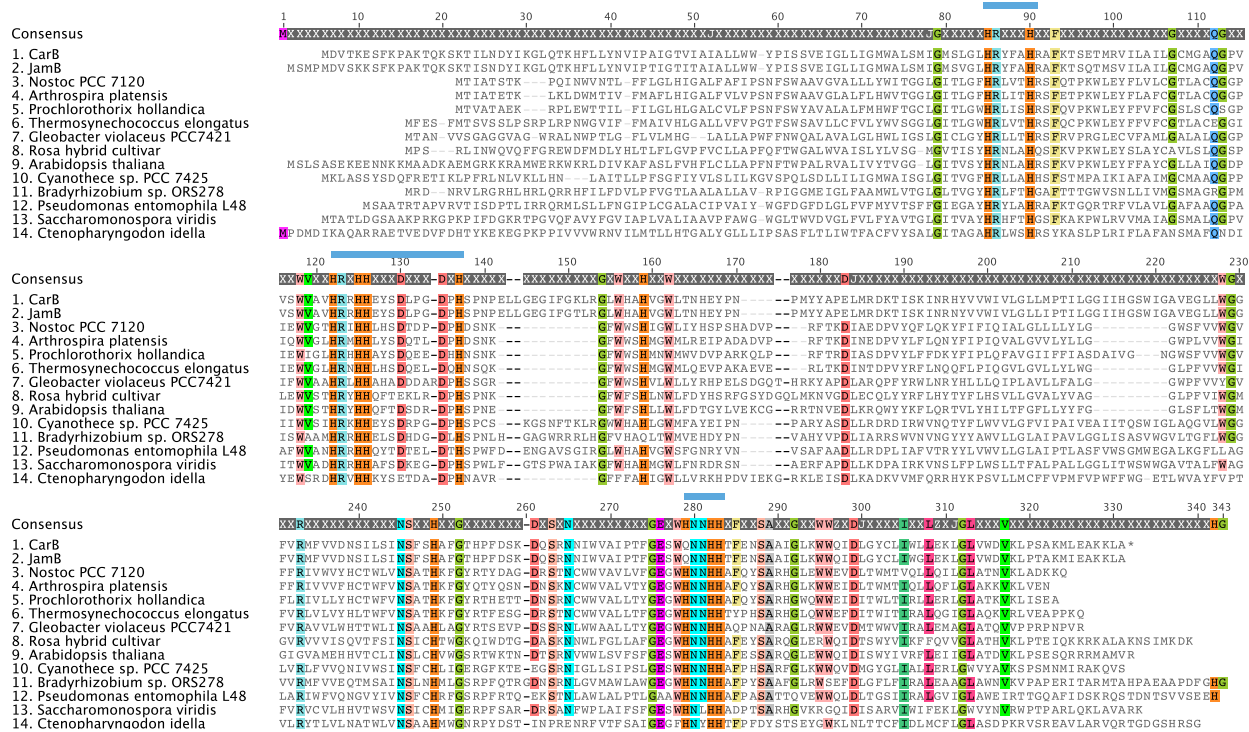


Figure 57. Alignment of bacterial and plant fatty acid desaturases. Blue bars indicate conserved histidine rich regions, responsible for iron coordination in the active site.

L. majuscula derived lipopeptides, carmabins A and B as well as a jamaicamides A, B and C have each been shown to make use of fatty acids with varying levels of unsaturation. These are likely formed by putative fatty acid desaturases (FAD) CarB and JamB present in respective biosynthetic gene clusters. CarB and JamB share very high sequence homology with DesC1 and DesC2 from *Spirulina platensis*, which are responsible for catalyzing double bond formation between C9 and C10 of stearic acid-derived, 18-carbon fatty acids, attached at the *sn*-1 and *sn*-2 positions of glycerol (88). These cyanobacterial desaturases belong to the membrane-bound fatty acid desaturase-like superfamily, which possesses characteristic hydrophobic regions that are proposed to be involved in membrane anchoring, as well as three conserved histidine-rich

clusters responsible for coordinating the two iron atoms in the protein active site (89) (Figure 57, 58).

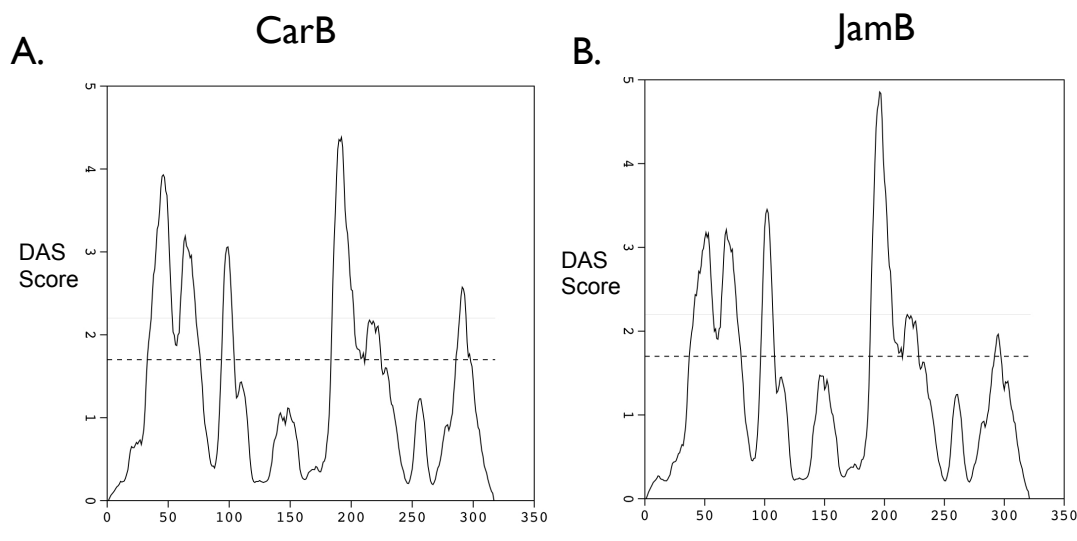


Figure 58. Density alignment score of fatty acid desaturases from carmabin and jamaicamide pathways. Transmembrane domains are predicted by DAS scores above 1.8. A. CarB; B. JamB.

4.4.1 Overexpression of CarB in a eukaryotic host

Investigation of this noncanonical desaturase initiated with *carB* DNA amplification from CarCos_L. The PCR product was cloned into a pET28 vector by traditional methods. The protein was overexpressed in BL21 (DE3) cells harboring the pRARE plasmid. Comparison of a time-course overexpression of the CarB compared with empty pET vector, showed no appreciation in abundance of a protein corresponding with the expected CarB molecular weight. Overexpression of membrane bound proteins using *E. coli* as a heterologous host are known to be challenging REF; In some cases this has been overcome through a combination of optimized fermentation conditions and the use of harsh purification protocols, often relying on chaotropic reagents and expensive detergents that do not ensure active enzyme.

In previous examinations of cyanobacterial fatty acid desaturases, yeast expression platforms have been used to successfully reconstitute enzymes (90). As such, synthetic *carB*, optimized for codon expression in yeast was prepared and subcloned into the appropriate vector for

overexpression in yeast. Recombinant protein overexpression was confirmed by time course expression study and analyzed by Western Blot using empty vector as a negative control.

Large-scale cultures of overexpressed proteins were harvested, and purified fatty acid desaturase was localized to the yeast microsomes (Figure 59).

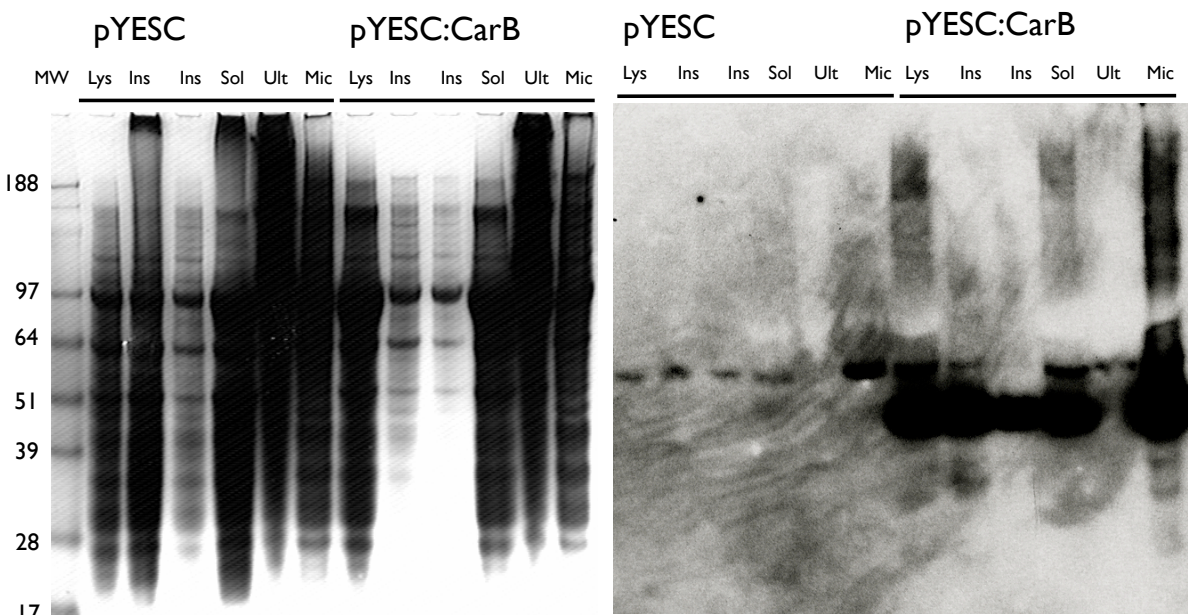


Figure 59. Microsome purification of CarB from yeast. Short chain fatty acid desaturase pYES:CarB was overexpressed and purified by in parallel with empty vector, pYESC. Asterisk denotes expected molecular weight ~43kDa. A. SDS-PAGE analysis of protein fractions; B. Western blot analysis of protein fractions. MW- molecular weight marker, Lys- lysate, Ins- insoluble fraction, Sol soluble fraction, Ult- supernatant fraction from ultracentrifuge, Mic- resuspended microsomes.

4.4.2 Fatty acid method development

CarB is proposed to catalyze the conversion of hexanoic acid to 5-hexynoic acid. *In vitro* characterization of enzymes oxidizing short and medium chain fatty acids is noticeably absent from the literature and required novel methods for reaction analysis. This is made challenging as neither the substrate nor product have a chemical functionality that is typically employed for biochemical analysis (i.e. UV or fluorescent chromophore, chemical handle for functionalization). Ideally, such a method would avoid radioactivity (i.e. ^{14}C), be amenable to tandem protein reaction analysis (i.e. CarA, B, C) and employ an available technology that enables determination of multiple enzymatic parameters.

4.4.2.1 Gas chromatography is a standard method for hydrocarbon analysis

Examination of short- and long-chain fatty acyl groups has been well achieved by coupled gas chromatography (GC) methods. Typically a GC is placed in-line to separate analytes, which are then introduced to the appropriate detector. Hydrocarbon analysis often employs flame ionization detectors (FID), which pyrolyze samples generating signal proportional to the abundance of organic atoms in the analyte. The relatively low boiling points of the C-6 acids make GC-FID an excellent candidate for analysis. GC chromatography method development proceeded with equal concentrations of hexanoic acid and 5-hexynoic acid prepared in ethyl acetate and the desired baseline separation of the compounds achieved (Figure 60).

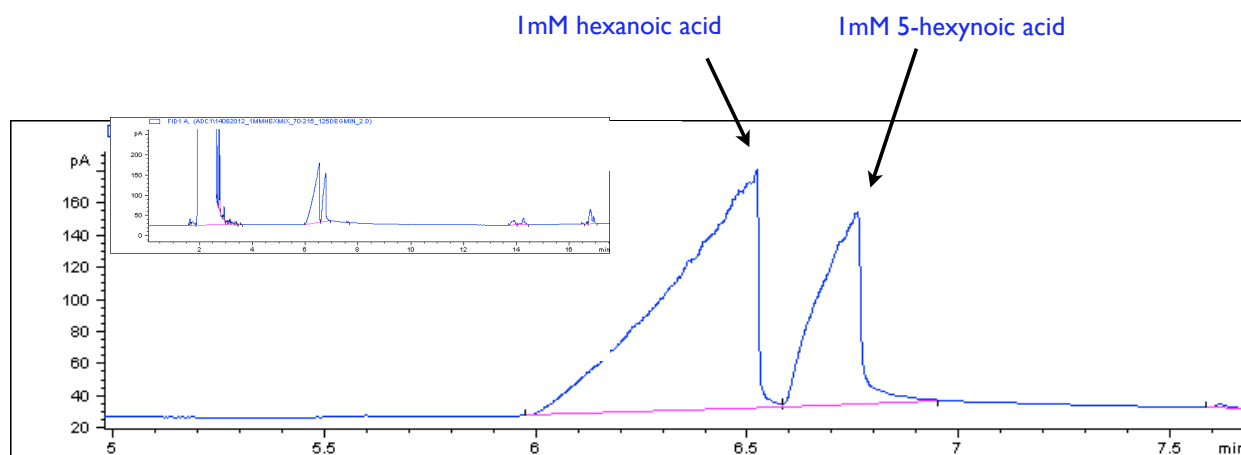


Figure 60. Short chain fatty acids separation by GC. Substrate, hexanoate, and product 5-hexynoate, of CarB reactions were prepared in ethyl acetate and analyzed by GC-FID. Inset: full chromatogram.

Interestingly, the fatty acids exhibited an elution profile with substantial peak fronting. This phenomenon was shown to be concentration dependent, with a uniform peak elution time.

To validate the method, substrate and product acids were prepared as diluted stocks, and added to the buffered reaction mixture lacking proteins. The samples were treated under standard assay conditions, extracted with ethyl acetate and analyzed by GC-FID. Unexpectedly, both substrate and product peaks showed substantially reduced signal following workup. Hexanoic acid showed 30% intensity of the control reaction, while the 5-hexynoic acid was less than 2% as intense as the control peak (Figure 61).

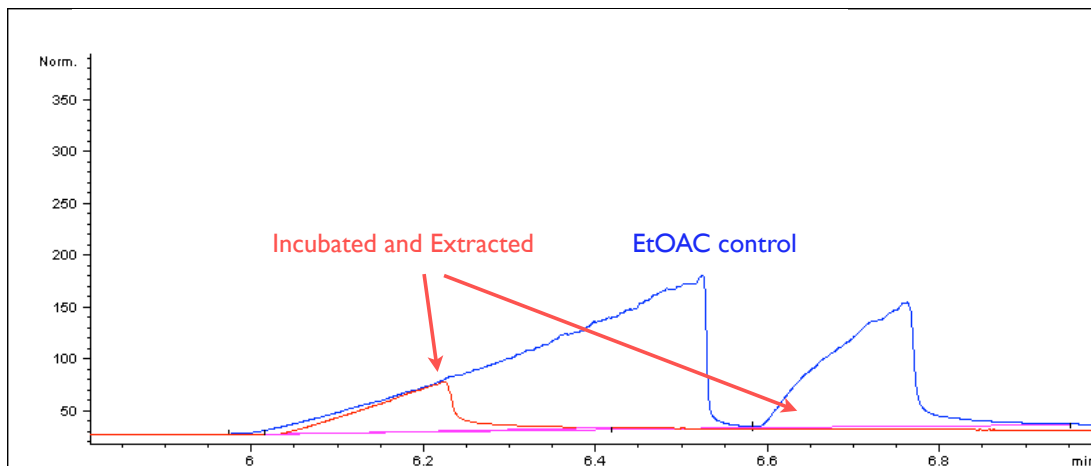


Figure 61. Fatty acids incubated in buffer show reduced signal. 1mM fatty acids were incubated and extracted with ethylacetate and analyzed by GC-FID

Though the reaction was prepared below the solubility limit of the free acids, protein amenable cosolvents were added to increase the stability of the samples in solution, however the addition of methanol, ethanol or dimethyl sulfoxide appeared to decrease the total signal intensity for both substrate and product (Figure 62).

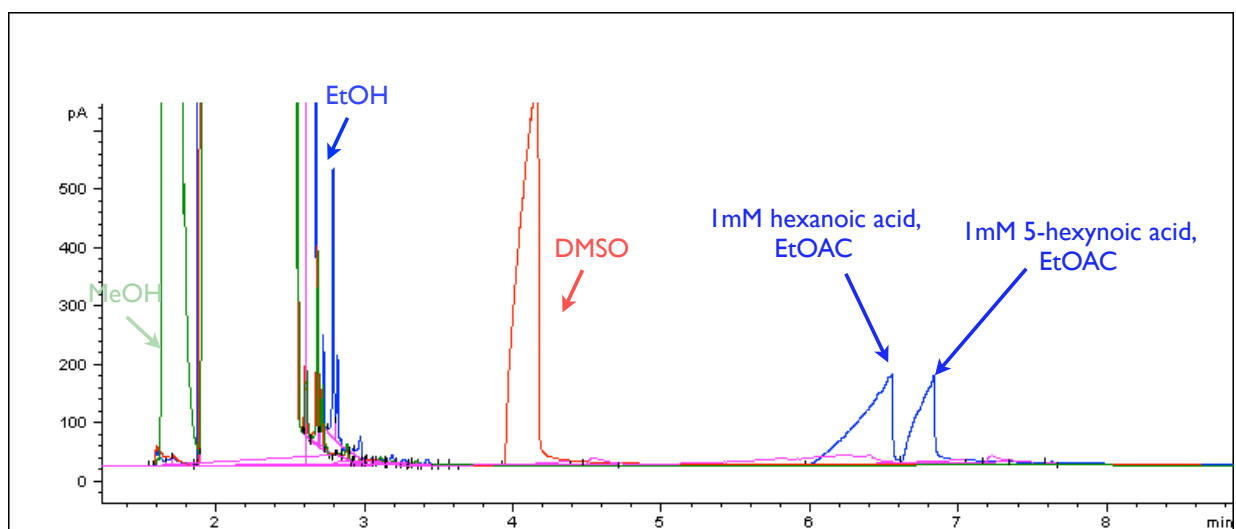


Figure 62. Cosolvents do not increase fatty acid detection by GC. Preparation of fatty acids in protein amenable in methanol, green trace, ethanol, blue trace, or dimethylsulfoxide, red trace, extracted and prepared for GC-FID

4.4.2.2 Headspace analysis of volatile compounds

The low solubility, high LogP, and difficulty detecting the substrate and product acids following incubation led to the hypothesis that the sample was volatilizing over the course of the reaction. GC methods forgoing extraction are typically employed when the analyte can be efficiently transferred to the gas phase and partition coefficient is sufficiently low for organic solvents. In order to test this, hexanoic acid and 5-hexynoic acid were prepared as before in gas-tight vials. Following incubation, samples were heated to 80° while shaking to transfer the maximal amount of sample to the gas phase. The headspace of the reaction vessel was sampled using a gas tight syringe, transferred to the GC-FID again no signal was detected.

4.4.2.3 Evaporative light scattering detection of fatty acids

The difficulty associated with compound detection by GC-FID, led to a reconsideration of the analytical method. Though, neither the substrate nor product possesses a chromophore for spectrophotometric detection, analysis could potentially be achieved through alternative optical methods. Evaporative light scattering detection (ELSD) is commonly employed in the analysis of non-UV absorbing lipids including cholesterols, cholesteresters and triacylglycerides. The detector is coupled to an HPLC in order to separate compound and samples are then detected through perturbations in optical turbidity of the nebulized HPLC eluent. To begin, samples containing either hexanoic acid or 5-hexynoic acid were prepared under reaction conditions, incubated, and injected directly without processing or workup. Substrate and product peaks were easily resolvable, with minor contaminants coming from impure standards also observed (Figure 63). This preliminary analytical method shows promise for characterization of this unexplored protein activity.

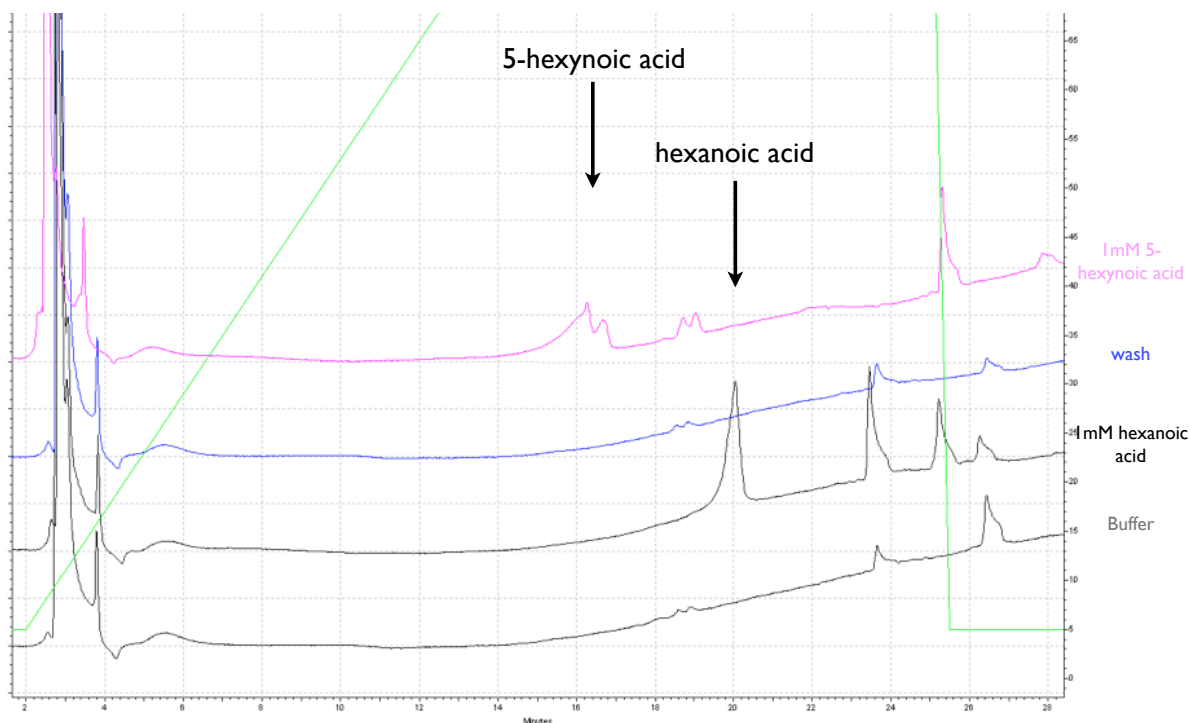


Figure 63. HPLC-coupled evaporative light scattering detection enables separation and visualization of short chain fatty acids. Samples denoted on right

4.5 Carmabin A possesses a terminal amide

Carmabin possesses a distinctive amide terminus that is shared by few other natural products, including the *Moorea producens* derived dragonamides (79, 82), macrocyclic antibiotic nosiheptide synthesized by *Streptomyces actuosus*, (81, 82) and myxathiazol derived from myxobacteria *Stigmatella aurantiaca* (83, 84). It remains unclear how the amide is formed, however isotopic feeding studies in *S. actuosus* indicated that the Nosiheptide amide is derived from serine. More recently, Weinig and coworkers demonstrated that the Myxathiazol amide arises as a result of C-terminal cleavage of a glycine added by an adenylation domain in the terminal biosynthetic module, MtaG. These experiments suggested that the carmabin amide may also be formed by an A domain possessing a noncanonical function.

4.5.1 CarI Terminal adenylation domain

Attempts to understand how the carmabin terminal amide is formed began with an investigation of the final synthetic module. Through bioinformatics studies, the CarI terminal domain was identified as a possible candidate for this reaction. Originally predicted to be a dehydrogenase,

reductase, or ornithine amino transferase, deep annotation of this domain, led to the renewed assignment as belonging to the LuxE superfamily of proteins. This class of enzyme, which also includes the NRPS adenylation domain, uses ATP to activate substrate acid and transfer to the downstream thiolation domain. A tree of representative dehydrogenase, reductase, ornithine aminotransferase and adenylation domains firmly places the final CarI domain within the adenylation clade (Figure 64). Thus, the nomenclature for this atypical domain is annotated as a terminal adenylation domain or TAD.

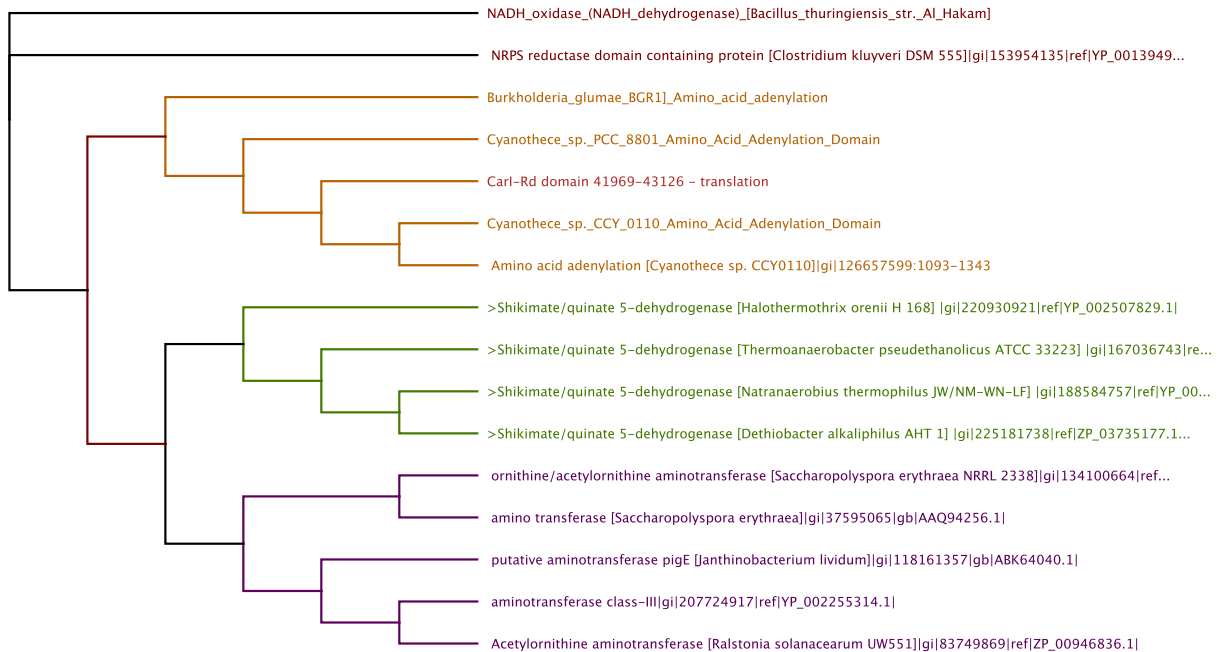


Figure 64. Cladogram for putative functions of the terminal domain from CarI.

Carl-TAD ATPase Activity

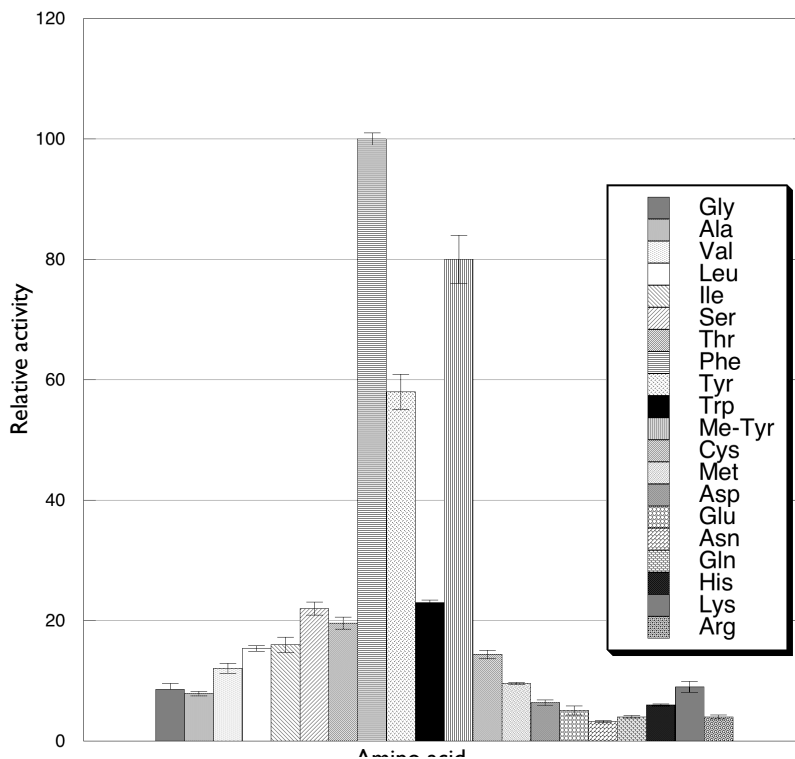


Figure 65. Pyrophosphate exchange assay to determine Carl terminal adenylation domain amino acid specificity. Radioactive count were normalized to phenylalanine for comparison.

To examine the activity of the TAD, the domain was cloned from CarCos_R using domain specific primers and subsequently cloned into pET28B. Transformed *E. coli* Rosetta (DE3) was used for protein overexpression. The enzyme was purified to homogeneity in a single step and stored. To determine the function of the terminal adenylation domain, the pyrophosphate exchange assay was employed. In this experiment the TAD catalyzes the forward reaction yielding an acyl-AMP. The high concentration of inorganic pyrophosphate drives the reverse reaction forming a mixture of radiolabeled and unlabeled ATP. The abundance of radioactive ATP is then correlated with substrate preference (Figure 65). The results of the pyrophosphate exchange assay suggest phenylalanine as the likely candidate for activation and use in transamination with a normalized $100 \pm 0.8\%$ activity, while the nearest activities are for O-methyl tyrosine and tyrosine with $79.9 \pm 4.4\%$ and $57.6 \pm 2\%$, respectively.

These results confirm the adenylation activity of the TAD. Based on the adenylation domain specificity profile it was expected that the TAD would be selective for an aromatic amino acid, which is in agreement with the results of the PPE assay.

4.5.2 CarI adenylation-methyltransferase-methyltransferase tridomain

The role of the duplicate adenylation domain architecture in CarI in forming the terminal amide of carmabin is unclear. To probe this unique enzyme system, adenylation-methyltransferase-methyltransferase (AMM) tridomain was cloned using standard molecular biology methods. The plasmid was transformed into BL21(DE3)-pLysS+pRARE, in order to minimize basal levels of the target gene.

Initially, overexpression of the protein proved challenging as the construct appeared to be toxic to the cells, with a complete loss of cell density within 2 hours following induction. The plasmid was used to transform BL21-AI, which has a single chromosomally integrated T7 RNA polymerase under an inducible arabinose promoter, which serves to prevent leaky transcription pre-IPTG induction and limit rates of mRNA transcription post-induction. This strategy proved successful to maintain cell viability during protein overexpression, however the majority of the target protein was insoluble.

The low proportion of the soluble protein may be the result of the improper folding as a result of the embedded methyltransferases within the adenylation domain. To increase the proper folding of the tridomain, TAKARA chaperone plasmids were screened for their ability to enhance the ratio of soluble protein and quantified by densitometry. The optimized overexpression was observed following the coexpression of DnaK-DnaJ-GrpE, increasing the soluble fraction by >400% (Figure 66).

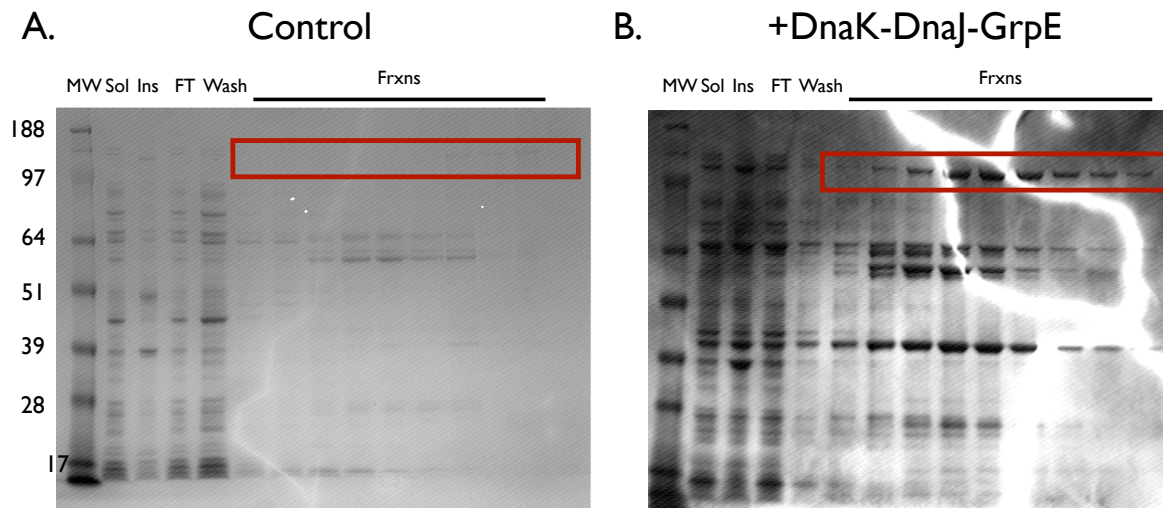


Figure 66. Overexpression of CarI AMM is dramatically increased with coexpression of protein chaperone. Expected MW= 144kDa. A. TB+ara B. TB+ara+DnaK-DnaJ-GrpE. MW- molecular weight marker, Sol- soluble fraction, Ins- insoluble fraction, FT- IMAC flowthrough, Wash- IMAC wash, FRXNS- IMAC stepwise fractions eluted with increasing concentration of imidazole

The addition of the non-denaturing, zwitterionic detergent CHAPSO enabled an increase in the total amount of the CarI_AMM protein, however a significant portion of contaminating proteins also co-purified (Figure 67).

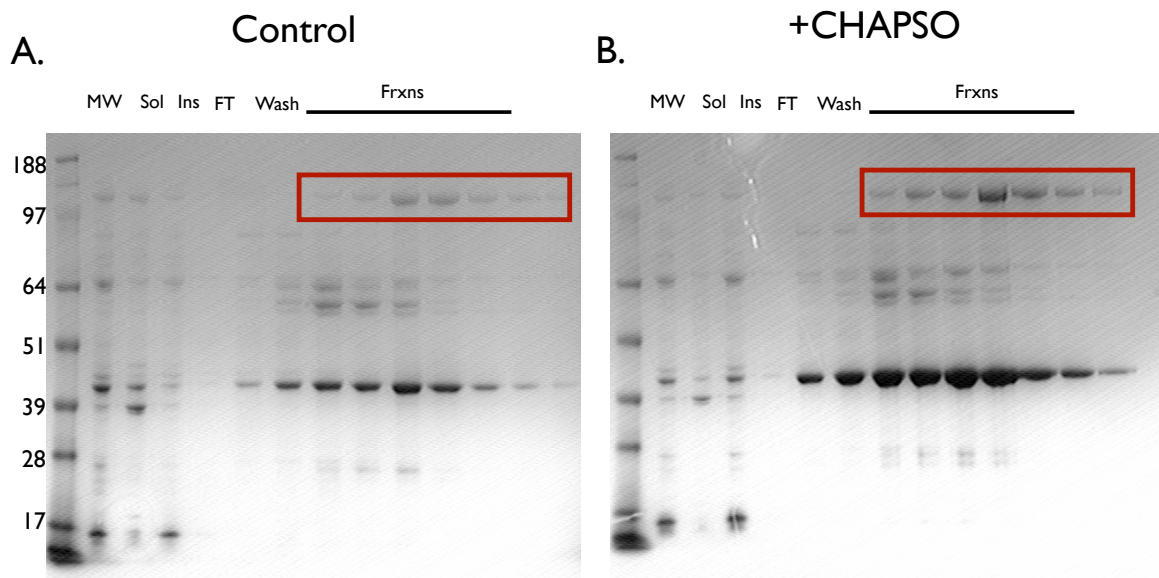


Figure 67. Overexpression of CarI AMM is slightly increased upon inclusion of CHAPSO detergent Expected MW= 144kDa. A. TB; B. TB+ CHAPSO. MW- molecular weight marker, Sol- soluble fraction, Ins- insoluble fraction, FT- IMAC flowthrough, Wash- IMAC wash, FRXNS- IMAC stepwise fractions eluted with increasing concentration of imidazole

This overexpression platform, while successful, was both time and reagent sensitive and a simplified method was sought. The addition of the chemical chaperone, TMAO has been used in the past to both solubilize and promote proper folding of proteins in *E. coli* overexpression systems. At low concentrations TMAO enabled the use of standard overexpression cell lines BL21DE3+pRARE, obviating the need for tight AI control (Figure 68). Increasing the concentration above 50mM however, did not further enhance protein solubility.

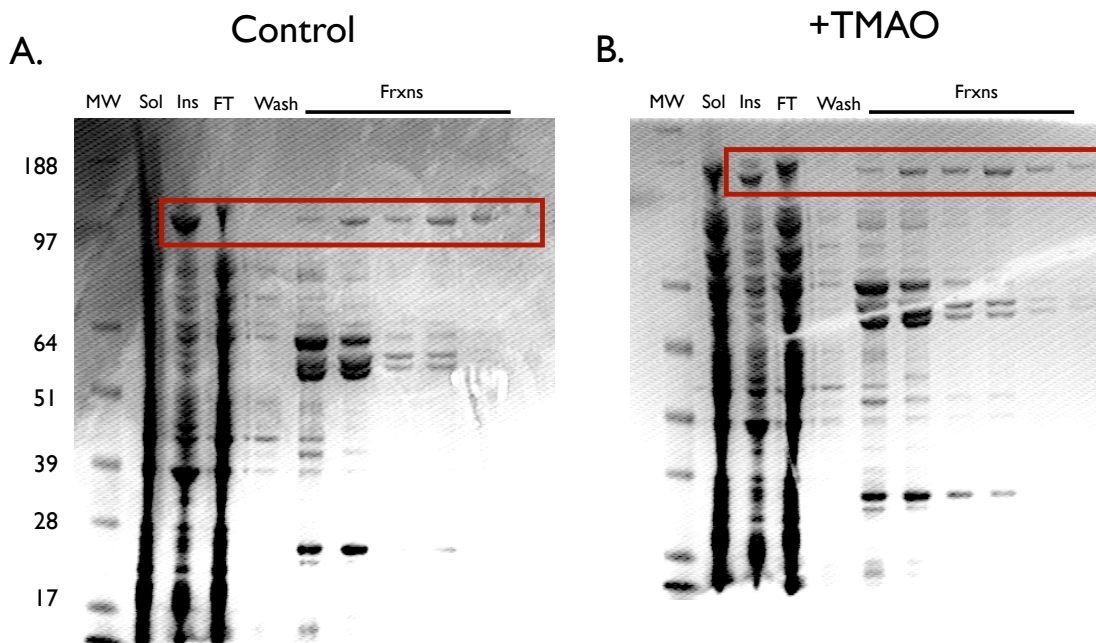


Figure 68. Overexpression of CarI AMM is slightly increased upon inclusion of chemical chaperone. Expected MW= 144kDa. A. TB; B. TB+ TMAO. MW- molecular weight marker, Sol- soluble fraction, Ins- insoluble fraction, FT- IMAC flowthrough, Wash- IMAC wash, FRXNS- IMAC stepwise fractions eluted with increasing concentration of imidazole

4.6 Discussion

Carmabin A is a structurally intriguing cyanobacterial natural product representative of a class of antimalarial lipopeptides. The initiating steps in carmabin biosynthesis leading to alkyne formation, putatively involve the activation and oxidation of hexanoic acid to 5-hexynoic acid via the acyl-CoA ligase (ACL) CarA, fatty acid desaturase (FAD) CarB and standalone carrier protein, CarC, though the mechanism is unclear. To date, all known cyanobacterial FADs have been shown to act on acyl chains that are bound to a head group as in glycerolipids, as activated CoAs, or bound to an acyl carrier protein (ACP) (88, 90-92). In vitro assays of membrane bound desaturases have demonstrated a ferredoxin dependence, suggesting a likely in vivo Red/Ox

partner, however detailed investigations into the activity and of FADs involved in natural products biosynthesis have yet to be conducted.

In the initial report describing biosynthesis of the alkyne-containing natural product jamaicamide A, Edwards et al. (27) successfully overexpressed and purified JamA, an acyl-CoA synthase (ACS). Once loaded, the FAD, JamB, was proposed to desaturate the fatty acid chain to form either an alkene or alkyne that are subsequently extended to form jamaicamides A-C. Contrary to the expected saturated hexanoic acid substrate, JamA showed an unexpectedly high preference for 5-hexenoic and 5-hexynoic acids via the pyrophosphate exchange assay. This lends support for a counter-canonical activity of JamB and CarB as unique FADs, with a proclivity toward utilizing free acids as substrates, suggesting desaturation occurs prior to loading onto the ACP. Interestingly, carmabin B possess a terminal methyl ketone, rather than the expected alkyne. The flexibility observed for JamA, to accept a variety of the fatty acid substrates, may be partially extended to the, CarA, as a possible explanation for the presence of this functional group.

Based on the structures, will also presumably see FAD domains as part of the biosynthetic machinery responsible for the antanapeptins, pitipeptolide, and wewakpeptins (21). Within cyanobacterial natural products the isolation of small molecule analogues with variable fatty acyl oxidation (e.g. antanapeptins A-C with alkyne, alkene and alkane functionalization, respectively) suggests that the downstream modules are tolerant of deviations from pathway colinearity and that early loading domains are substrate flexible, in agreement with the semi-promiscuous activity observed for JamA.

The termination events leading to amide formation in carmabin A similarly remain elusive. Bioinformatic analysis and biochemical characterization of the CarI terminal domain (TAD) via the pyrophosphate exchange assay identified this protein as an adenyating enzyme plausible candidate for catalyzing transamination. Potentially, release involves the addition of second phenylalanine, activated by the CarI TAD, of which only the amine remains in the mature molecule. Isotopic feeding studies using ¹⁵N-labeled amino acids could be used to decisively determine the origins of the nitrogen.

While terminal amides in natural products are not common, myxobacterial metabolites have demonstrated an unusual abundance, of which the biosynthetic origins have begun to be characterized. Recently, the myxathiazol amide was discovered to be the result of a complete NRPS module, MtaG, using an N-acetylcysteamine substrate surrogate (84). Here, the A- and C-domains catalyzed activation and condensation of glycine onto the growing chain, attached to the downstream peptidyl carrier protein (PCP). An embedded flavin monooxygenase catalyzed oxidative chain release, resulting in the amide-containing NP and a PCP-bound 2-oxoacetate. In order to regenerate the holo-PCP, the adjacent Type-II thioesterase (TE) hydrolyzed the thioester yielding the reduced phosphopantetheine thiol and free glycolic acid. This type of housekeeping ensures the viability of the terminal module for multiple rounds of biosynthetic catalysis. Interestingly, characterization of the closely related ester-homologue, melithiazol led to the identification of two additional proteins in that cluster, MelJ and MelK, that catalyzed amidohydrolase and methyl transferase activities, respectively (83). This finding indicates that amide offloading of natural products may be more a more common strategy during biosynthesis than previously understood, based exclusively on the structure of the isolated compounds.

Deep annotation of the carmabin gene cluster resulted in the identification of a previously unrecognized Type-II TE. The presence of this terminal protein in the carmabin pathway supports the hypothesis that the carmabin A amide is amino acid derived in a fashion potentially similar to Myxothiazol. The standalone TE may also play a role as a holding domain, similar to what has been observed in iterative NRPS gramicidin S pathway (93). Hoyer *et al.* demonstrated that during biosynthesis, one half of the dimeric pentapeptide Phe-Pro-Val-Orn-Leu, was fully elongated, and transferred to the GrsB-TE. The second pentapeptide was again generated by the NRPS and attached to the terminal PCP. Remarkably the TE catalyzed both dimerization of the peptides, as well as subsequent cyclization.

Unlike myxathiazol biosynthesis, the absence of a dedicated transamidating module in the carmabin cluster compels a further investigation into the potential function of the terminal CarI domain. Recently, bifunctional adenylation domains have been identified in a number of natural product pathways capable of catalyzing not only amino acid activation but also peptide condensation. In the biosynthesis of the aminoglycoside streptothricin, a standalone domain, Orf

19, was demonstrated to possess both the A- and C-domain like activities. This 59 kDa protein, activated β -lysine as the amino acyl-AMP, as well as catalyzed subsequent head-to-tail condensation with a carrier protein bound β -lysine, loaded by a separate standalone adenylation domain, Orf 5. Remarkably, Orf 19 was able to catalyze seven iterative rounds of tandem activation and condensation, in agreement with the longest β -lysine chain length found on streptothricin X.

Similarly, biosynthesis of the aminocoumarin antibiotics novobiocin and coumermycin A₁, involve a bifunctional amide synthetase(94, 95). NovL and CouL were both shown to catalyze an initial adenylation step with 3-dimethylallyl-4-hydroxybenzoic acid and 3-methyl-1-pyrrole-2,4-dicarboxylic acid, as substrates respectively. Following activation, the acyl-AMPs were condensed with the core 3-amino-8-methyl-4,7-dihydroxycoumarin ring. Interestingly, CouL was also responsible for a second round of activation and condensation with the carboxylic acid at the C4 position and an additional dihydroxycoumarin ring.

With the groundwork in place, investigating alkyne formation and amide production in carmabin A are rapidly taking shape. Understanding the precise mechanism of the unusual initiation and termination events will guide future examination of marine cyanobacterial biosynthesis.

4.7 Materials and methods

Bacterial strains, media and culture conditions.

Escherichia coli DH5 α subcloning efficiency cells (Invitrogen) were used as axenic stocks for the preparation of electrocompetent cells for DNA propagation and maintenance. *Escherichia coli* BL21-pLysS and BL21-AI (Novagen) harboring the pRARE plasmid (spectinomycin resistance exchanged for chloramphenicol by Dr. W. Clay Brown, University of Michigan) were used for general protein overexpression. All transformations were achieved via electroporation (BioRad, Gene Pulser Xcell), with 2.5-500 ng DNA from ligation reaction mixture or as intact plasmid, recovered with SOB, allowed to outgrow for 60 minutes prior to plating on lysogeny broth-Miller-agar plates (LB-Miller) supplemented with the appropriate antibiotic (ampicillin- 50 μ g/mL, kanamycin- 50 μ g/mL).

Protein overexpression using *E. coli* as a host organism

Fresh transformants harboring the plasmid of interest were used to inoculate a culture of lysogeny broth-Miller (LB-Miller). Overnight (~12-18 hours) cultures were used to subinoculate (1:100 (v/v)) large-scale (500-1000 mL) cultures of LB-Miller for the TAD and either 2XYT or Terrific Broth (TB) for the A-MT-MT tridomain, supplemented with appropriate antibiotics for protein overexpression.

For the TAD and large-scales cultures were grown at 37° while shaking at 160 RPM until reaching an OD₆₀₀ 0.8-1.0. Flasks were allowed to cool at room temperature for 60 minutes. Protein overexpression was induced with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM. Flasks were returned to an 18°C shaker and continued to grow for an additional 12-14 hours. A-MT-MT tridomain were treated similarly with the addition of 50 mM TMAO to the culture media. A-MT-MT producing cells grown in BL21-AI were grown at 37°C until an OD of 0.4, at which point arabinose was added to a final concentration of 0.2 mg/mL. The culture cooled passively in the shaker at a temperature of 24°C until an OD of 1.0 was reached. Overexpression was initiated following the addition of 0.1mM IPTG to the culture, which was allowed to grow at 15°C for an additional 10-16 hours. Cells were harvested by centrifugation at (5000xg, 10 minutes, 4°), transferred to 50 mL Falcon tubes and cell pellets were flash frozen in liquid nitrogen and stored at -80° until purification.

Protein purification from an *E. coli* heterologous host

All steps during the protein purifications were performed at 4°. Cell pellets were resuspended in 2x cell volume (minimum 25mL) 5% Buffer B, and incubated with freshly prepared lysozyme (0.2 mg/mL final, prepared in MilliQ H₂O), freshly prepared PMSF (2 mM final, prepared in DMSO) and 1 *cOmplete* protease inhibitor cocktail (Roche) tablet for 60 minutes while rocking. Cells were lysed by use of a recently tuned ultrasonic dismembrator for 10 cycles of 1 minute, with pulses of 5 seconds on, followed by 5 seconds off, and intensities ranging from 7-8. All cycles were carried with chilled beakers in packed ice bucket. Cellular debris was removed by centrifugation at 40000-50000xg for 30 minutes and the supernatant was transferred to a fresh

tube; residual insoluble material was removed by filtration (syringe driven 0.45 μm filter) or a second round of centrifugation.

Immobilized metal affinity chromatography

Clarified lysate was applied to a 5 mL HiTrap IMAC column (GE) charged with Ni^{2+} , and equilibrated with 5% Buffer B prepared with or without CHAPSO mounted on an AKTA FPLC (GE). The resin was washed extensively with 25 column volumes of 10% buffer B to remove any nonspecifically bound proteins. The target protein was eluted with a linear gradient from 15-100% Buffer B. Peak fractions as determined by gradient SDS-PAGE ((4-12%) (Novex NuPage) run in MOPS buffer at 200 mV for 30 minutes) were pooled and concentrated using Amicon ultra centrifugal concentrators (Millipore, 10 kDa, 50 kDa), equilibrated with Buffer F until a volume of 2.5 mL was reached. Proteins were buffer exchanged with a PD10 washed and equilibrated with Buffer F and eluted with 3.5 mL of Buffer F. Proteins were further concentrated to less than 100-500 μL , aliquoted, flash frozen in liquid nitrogen and stored at -80° . Protein concentrations were determined using the bicinchoninic acid assay (BCA Pierce) according to the manufacturers guidelines, using a BSA standard curve and read in 96 well format.

Yeast protein overexpression of the fatty acid desaturase

Synthetically prepared CarB (GeneArt) was subcloned from pMA vector into the pYESC yeast expression vector (Life Technologies) according standard molecular biology practices. The uracil auxotrophic INVSC-1 yeast strain was transformed with the pYES-CarB according to the manufacturer's guidelines. Protein overexpression and yeast microsome preparation were carried out in concordance with protocols outlined by the manufacturer with no modification. Detection of the overexpressed yeast protein was carried out by Western blot and using the Anti-HisG-HRP antibody (Life Technologies) and visualized using the ECL Prime Western body kit (GE Healthcare).

Short chain fatty acid analysis by GC-FID and RP-HPLC-ELSD

Method development for GC-FID analysis of fatty acids was guided by the work described in Das *et al.* (96). Modifications to the temperature gradient for extracted samples were optimized

to ensure baseline separation of the C-6 acids. Preparation of the reaction mixture for fatty acid detection by ELSD was the same as above, excluding the extraction with ethyl acetate. Samples were injected directly onto an ACE AQ column (ACE 3 μm , 100 mmx 3.0 mm) equilibrated in 10% CH_3CN mounted on a Beckman Coulter HPLC. A gradient of 10-95% CH_3CN over 20 minutes, followed by a 5 minute wash at 95% CH_3CN and reequilibration at 10% $\text{CH}_3\text{CN}+0.1\%\text{FA}$ for 10 minutes. Column elution was directed to an ELSD (Alltech) with N_2 nebulizing gas running at 1 l/min.

Pyrophosphate exchange characterization of the TAD

Activity assay characterizing the domain activity the terminal adenylyating enzyme was carried out using a protocol adapted from (97, 98). Briefly in a 100 μL reaction (50 mM Tris (pH= 7.2) 100 mM NaCl, 1 mM MgCl_2 , 10% (v/v) glycerol, 10 mM ATP, 1 mM DTT ^{32}P labeled pyrophosphate (100 μM , 2 $\mu\text{Ci}/\mu\text{Mol}$) 500 μM cold pyrophosphate) 5 mM amino acid was incubated with 10 μM TAD for 60 minutes at 24°C. The reactions were quenched with the addition of 10% (v/v) ice-cold trichloroacetic acid (TCA) and followed by 500 μL of a Norit suspension (2.0% (w/v) activated charcoal. The activated charcoal was pelleted by and washed thrice with 1 mL of 100 μM pyrophosphate and 10% TCA in MilliQ H_2O . The charcoal was resuspended in 500 μL MilliQ H_2O and 3.5 mL scintillation solution and percent bound determined by scintillation counting.

Acknowledgements: This work is the combined effort of the Sherman and Gerwick labs. Dr. Junyong Jia (JJ), Dr. Adam C. Jones (ACJ), and Dr. Lena Gerwick prepared *Moorea producens* cosmid library and sequenced the cyanobacterial genome. Cosmid screening and preliminary pathway identification was done by JJ and ACJ. Deep pathway annotation, and all molecular biology and biochemistry was performed by Eli Benchell Eisman with support from Dr. Liangcai Gu and Dr. Bryan Yestrepky.

Chapter 5

Future directions

5.1 Megasyntase enzymology

Exploring polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) global enzyme strategy is a rapidly expanding field of study, relying increasingly on advanced techniques and methodologies to parse the complex and multistep biosynthetic process (**10, 11**). Evolutionarily, PKSs are presumed to derive from enzymes involved in fatty acid metabolism (**99**). Bacterial Type II fatty acid synthases (FAS-II) are composed of discrete domains which complex to catalyze formation of fully saturated acyl chains. FAS domain shuffling and exclusion likely led to ancestral PKS pathways with architecture similar to the *trans*-acting Type-II PKS systems, while the megaenzymes that we observe for Type-I PKSs resulted from gene fusion of co-regulated clusters.

While the overall biosynthetic product is similar between FAS and PKS systems, the mechanism by which the substrate, bound to the acyl carrier protein (ACP) within a given module or as part of the *trans*-protein-complex, discriminately interacts with individual catalytic domain remains elusive. The unusual mixed *cis*- and *trans*- acting β -branching systems(37), as exemplified by CurA-CurF, further complicate this process requirements by requiring a tandem carrier proteins array to engage with domains nonsequentially during single catalytic cycle (i.e. substrate bound to the CurA ACP_{I+II+III+Cd} must first undergo extension carried out by CurA PKS domains, 3-hydroxy-3-methylglutaryl formation is catalyzed by the standalone HMG-CoA synthase, reengagement with the intramodular Halogenase for chlorination to occur, followed by dehydration by the standalone enoyl-CoA hydratase, decarboxylation by a second enoyl-CoA

hydratase as part of the downstream module and subsequent cyclization yielding the 2-methyl-cyclopropyl moiety) (40). Moreover the role of the C-terminal flanking domain (Cd) in mediating specific contacts as a means of facilitating enhanced biosynthetic flux through that system is incredibly unique(38).

5.1.1 Type-I PKS modular strategy

Inroads into understanding the forces that govern such protein-protein interactions are beginning to be made. Notably, the recent kinetic characterization of modular loading events by Bonnet and Rath (100) provide evidence for an effect of the substrate and ACP isoform on the rate of catalysis by the acyl transferase (AT). By monitoring free CoA production, under steady-state conditions, wild type PikAIV was demonstrated to catalyze processing of the native substrate, methylmalonyl-CoA, 1.4-fold faster in the presence of a phosphopantetheinylated, holo-ACP, rather than non-posttranslationally modified, apo-ACP, demonstrating the significance in competent interaction between domains in the priming event during in the first step in PKS function. Interestingly, for the non-native substrate malonyl-CoA, the AT appeared 5.17-9.36-fold more active for the holo- and apo- ACPs, respectively. This apparent increase in rate can be understood to be the result of the evolved discrimination of the AT and ACP domains for the correct extender unit, indicating the marked role that the substrate and plays in driving PKS flux.

These conclusions are further supported by the characterization of the PikAIII module by Whicher and Dutta (unpublished data), wherein cryo-electron microscopy revealed structural rearrangements and remarkably precise localization of the ACP over the course of a single round of modular catalysis in a principally substrate driven mode. In the presence of keto synthase (KS) loaded pentaketide, but absence of the extender unit, holo-ACP resided exclusively at the AT, pausing until loading can take place. Withholding chain elongation intermediate, but providing the extender unit methylmalonyl-CoA led to the ACP occupying space below the internal KS active site tunnel. Addition of both the pentaketide and methylmalonyl-CoA, yielded the extended β -keto-hexaketide and a complete repositioning of the ACP to the ketoreductase (KR) domain. Supplementation of the reaction with NADPH, led to β -alcohol formation and unexpectedly the ACP appeared to eject from within the PKS dimer active site arch, in position to engage with the KS of the downstream PikAIV module and via the well

established donor-acceptor docking interactions. The unexpected spatial localization of the ACP in response to the loaded substrate clearly demonstrates a process by which the noncatalytic ACP is involved in recognition and transposition of the nascent intermediate in order to engage with individual catalytic domains at explicit moments during catalysis.

5.1.2 Iterative Type-I PKS enzymology

These studies demonstrate that within a module, during a single PKS cycle, the sequence of protein-protein interactions is substrate determined. An expansion of these works characterizing the interdomain recognition of substrate within Type-I systems can be expanded to probe the intramodular interactions involved in iterative PKS (IPKS) metabolism. Typically observed in fungi, a single module consists of the core PKS domains as well as the full complement reducing enzymes (KR, dehydratase, (DH) and enoyl reductase (ER)), and is wholly responsible for small molecule scaffold assembly (*101*). In the biosynthesis of the fungal metabolite lovastatin, LovB, an eight-domain module (KS-AT-DH-MT-ER⁰-KR-ACP-C) and stand alone ER, LovC, catalyze the condensation and nine malonyl-CoAs generating the natural product core dihydromonacolin L (*102*). Totally unknown are the forces that constrain individual proteins to be either excluded or included during modular turnover, however recent evidence suggests that this may also be substrate guided; this is especially apparent for the α -methylation event during the fourth round of LovB catalysis. In the absence of productive methyl transferase interaction, experimentally by withholding S-adenosyl-methionine from the reaction mixture, Ma *et al.* demonstrated that the IPKS module is still able to process substrate, however biosynthesis leads to undesired pyrone formation and premature offloading.

In cytochalasan biosynthesis, the IPKS modules are further expanded to incorporate the three core NRPS domains (condensation (C), adenylation (A), peptidyl carrier protein (PCP)) and reductase (Red), C-terminal of the ACP. Counterintuitively, the natural products generated by these pathways have only a single amino acid as part of the final structure. The shape adopted by this hybrid IPKS-NRPS in protein-space and mode by which the iterative PKS activity remains disconnected from NRPS function is totally unknown. The strategy used by these more exotic PKS systems to delineate intermediate processing and domain incorporation in generating the desired product is clearly in response to ACP-bound substrate.

5.1.3 Acyl carrier proteins as active participants during small molecule assembly

Understanding the forces that govern PKS turnover will provide insight into the discrimination of CurM and *Synechococcus* olefin synthase to offload the metabolite as the alkene, as well as the mode by which β -branching proteins interact. The role of the carrier protein in guiding interaction has begun to be investigated in a number of Type-I systems, however with little consensus. Evans and coworkers (**103**) have observed conformational dynamics that take place during the actinorhodin ACP transition from apo- to holo-, with a corresponding decrease in affinity for the phosphopantetheinyl (PPant) transferase, FAS-holo-synthase (ACPS). Conversely, Busche *et al.* (**53**) only identified minor structural changes of the CurA ACP triplet upon addition of the PPant. In studying fatty acid metabolism, Ploskon *et al.* (**104**) observed significant interactions of the proximal portion of the acyl substrate and a hydrophobic binding cleft generated by α HIII and α HIV of the *Streptomyces coelicolor* FAS ACP, however no such sequestering was observed for the homologous human FAS proteins (**105**). A comprehensive appreciation of the substrate-ACP and ACP-catalytic domain associations are critical for reaching the long-term goal of secondary metabolite pathway engineering. Similar to PKS systems, protein-bound substrates within NRPSs must engage with catalytic domains in a precise manner, though a less thorough treatment exists in the literature (**106**). By minimizing the judicious interactions, which limit productive modular function, nonnative substrates can be successfully incorporated into the final compound. Advanced characterization of the specific interdomain contacts within PKS, IPKS, NRPS, mixed systems and pathways including β -branching cassettes will certainly be a component of the next stage in megasynthase enzymology.

5.2 Chain elongation intermediates as pseudo-activity based protein-profiling probes

A major gap in the natural product sciences, is the disconnect between the isolation of small molecules and identification of the responsible biosynthetic gene cluster. As next generation sequencing technologies become accessible and bioinformatic tools become progressively refined, the frequency and ease with which whole genomes are screened for biosynthetic pathways is rapidly increasing (**11**). Practically, this can still be challenging, as there are often silent or cryptic gene clusters with no corresponding small molecule, as well as secondary

metabolites that may deviate from the product as predicted by the biosynthetic protein template, both of which obfuscate analysis. Experimentally Rath *et al.* (16) partially addressed this limitation by simultaneously characterizing the metagenome, metaproteome and metametabolome of a marine invertebrate and associated microbiome known to produce the chemotherapeutic ET-743. While both time and resource intensive, this method demonstrated the utility of a contemporaneous examination of a single system using multiple biological metrics as a means validating the assignment of a small molecule to biosynthetic machine; however, this technique was unable to ascribe complete domain function for key metabolic steps.

The terminal unsaturation present on carmabin A, has incredible implications for use in deciphering natural product biosynthesis. As a chemical handle, alkynes have been widely exploited for a rapid and specific functionalization via bioorthogonal chemistry (107, 108). Acting as a pseudo-activity based protein profiling (pABPP) probe (109-114), alkyne-containing pathway intermediates, could be derivatized using “Clickable” techniques. One such scenario envisions the addition of azide-linked biotin to lysed cyanobacterial culture during peak carmabin production, which undergo the expected [3+2] cycloaddition. As the nascent chain remains chemically tethered to the module during catalysis via the carrier protein thioester bond, all biosynthetic proteins could be enriched for via streptavidin pulldown. Peptide mapping by bottom-up mass spectrometry (MS) could then be used to determine the protein identity and the associated biosynthetic intermediate could simultaneously be determined by tandem MS. This single-step enrichment of both the metabolite and corresponding enzymatic machinery from crude cell mass vastly simplifies the pathway identification workflow and would provide direct evidence for the order of catalytic steps.

Elements of this proposal have already been shown successful in the literature. In a proof of principle study examining NPRS interaction Hur *et al.* (115) demonstrated that chemical crosslinking can take place between cognate modules upon loading PCP domains with either an alkyne or azide appended to a PPant derivative. Meier *et al.* (116) have taken advantage of the presence of the ubiquitous phosphopantetheine group from enriched biological material as a diagnostic mass spectrometry tool in order to confirm and sequence-identify the presence of a whole modules from crude sample. Similarly, Yap and Kostiuk *et al.* (117) have shown that

proteins post-translationally modified with and alkynyl fatty acids successfully undergo triazole formation with a biotin-linked azide, under *in vivo* and *in vitro* conditions. Moreover the protein conjugates could be easily enriched from HeLa cells and identified by streptavidin pulldown and bottom-up MS.

Through this combinatorial chemical-biology approach and utilizing recent advances in mass spectral networking(118, 119), families of natural products, along with the corresponding intermediates can be rapidly identified. Furthermore, the enzymatic machinery responsible for the biosynthesis of such small molecules could also be identified in a single step. The experimental workflow attempts to broaden the use of natural product intermediates as pABPPs, however this is met by the obvious limitation of the finite number of small molecules with a terminal unsaturation. Common bioconjugate reagents such as isocyanates, which are both hydroxyl and amine reactive, would be unable to distinguish between small molecule and biomolecule and would be an unreasonable alternative. An option to include an initial protein chromatographic step would help bias toward the larger megasynthases and could be followed by parallel examination by multiple reactive probes could. Ultimately the technique outlined above will be used to advance a unified method MS-based, multi-omic approach to natural product discovery.

References

1. Harvey, A. (2008) Natural products in drug discovery, *Drug Discovery Today* 13, 894-901.
2. Butler, M. S. (2005) Natural products to drugs: natural product derived compounds in clinical trials, *Natural product reports* 22, 162.
3. Carter, G. T. (2011) Natural products and Pharma 2011: Strategic changes spur new opportunities, *Natural product reports* 28, 1783.
4. Gray, A. I., Igoli, J. O., and Edrada-Ebel, R. (2012) Natural Products Isolation in Modern Drug Discovery Programs, In *link.springer.com*, pp 515-534, Humana Press, Totowa, NJ.
5. Newman, D. J., and Cragg, G. M. (2012) Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010, *Journal of Natural Products* 75, 311-335.
6. Li, J. W. H., and Vederas, J. C. (2009) Drug discovery and natural products: end of an era or an endless frontier?, *Science* 325, 161-165.
7. Cragg, G. M., and Newman, D. J. (2013) Natural products: A continuing source of novel drug leads, *Biochimica et Biophysica Acta (BBA) - General Subjects* 1830, 3670-3695.
8. Fischbach, M. A., and Walsh, C. T. (2006) Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms, *Chemical Reviews* 106, 3468-3496.
9. Konz, D., and Marahiel, M. A. (1999) How do peptide synthetases generate structural diversity?, *Chemistry & Biology* 6, R39-48.
10. Schwarzer, D., and Marahiel, M. A. (2001) Multimodular biocatalysts for natural product assembly, *Naturwissenschaften* 88, 93-101.
11. Walsh, C. T., and Fischbach, M. A. (2010) Natural Products Version 2.0: Connecting Genes to Molecules, *Journal of the American Chemical Society* 132, 2469-2493.
12. Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., and Gerwick, W. H. (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*., *Journal of natural*
13. Du, L., and Lou, L. (2010) PKS and NRPS release mechanisms, *Natural product reports* 27, 255.
14. Kittendorf, J. D., Beck, B. J., Buchholz, T. J., Seufert, W., and Sherman, D. H. (2007) Interrogating the Molecular Basis for Multiple Macrolactone Ring Formation by the Pikromycin Polyketide Synthase, *Chemistry & Biology* 14, 944-954.

15. Wuest, W. M., Krahn, D., Kaiser, M., and Walsh, C. T. (2011) Enzymatic Timing and Tailoring of Macrolactamization in Syringolin Biosynthesis, *Organic Letters* 13, 4518-4521.
16. Rath, C. M., Janto, B., Earl, J., Ahmed, A., Hu, F. Z., Hiller, L., Dahlgren, M., Kreft, R., Yu, F., Wolff, J. J., Kweon, H. K., Christiansen, M. A., Håkansson, K., Williams, R. M., Ehrlich, G. D., and Sherman, D. H. (2011) Meta-omic Characterization of the Marine Invertebrate Microbial Consortium That Produces the Chemotherapeutic Natural Product ET-743, *ACS Chemical Biology* 6, 1244-1256.
17. Silakowski, B., Nordsiek, G., Kunze, B., Blöcker, H., and Müller, R. (2001) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15, *Chemistry & Biology* 8, 59-69.
18. Paul, V. J., and Ritson-Williams, R. (2008) Marine chemical ecology, *Natural product reports* 25, 662.
19. Waters, A. L., Hill, R. T., Place, A. R., and Hamann, M. T. (2010) The expanding role of marine microbes in pharmaceutical development, *Current Opinion in Biotechnology*, 1-7.
20. Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., and Prinsep, M. R. (2013) Marine natural products., *Natural product reports* 30, 237-323.
21. Tan, L. T. (2007) Bioactive natural products from marine cyanobacteria for drug discovery, *Phytochemistry* 68, 954-979.
22. Engene, N., Choi, H., Esquenazi, E., Rottacker, E. C., Ellisman, M. H., Dorrestein, P. C., and Gerwick, W. H. (2011) Underestimated biodiversity as a major explanation for the perceived rich secondary metabolite capacity of the cyanobacterial genus *Lyngbya*, *Environmental Microbiology* 13, 1601-1610.
23. Engene, N., Rottacker, E. C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M. H., Komárek, J., and Gerwick, W. H. (2012) *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites, *International Journal of Systematic and Evolutionary Microbiology* 62, 1171-1178.
24. Chang, Z., Flatt, P., Gerwick, W. H., Nguyen, V.-A., Willis, C. L., and Sherman, D. H. (2002) The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit., *Gene* 296, 235-247.
25. Edwards, D. J., and Gerwick, W. H. (2004) Lyngbyatoxin Biosynthesis: Sequence of Biosynthetic Gene Cluster and Identification of a Novel Aromatic Prenyltransferase, *Journal of the American Chemical Society* 126, 11432-11433.
26. Ramaswamy, A. V., Sorrels, C. M., and Gerwick, W. H. (2007) Cloning and Biochemical Characterization of the Hectochlorin Biosynthetic Gene Cluster from the Marine Cyanobacterium *Lyngbya majuscula*, *Journal of Natural Products* 70, 1977-1986.
27. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., and Gerwick, W. H. (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*., *Chemistry & Biology* 11, 817-833.
28. Hooper, G. J., Orjala, J., Schatzman, R. C., and Gerwick, W. H. (1998) Carmabins A and B, New Lipopeptides from the Caribbean Cyanobacterium *Lyngbya majuscula*, *Journal of Natural Products* 61, 529-533.

29. Gerwick, W. H., Proteau, P. J., and Nagle, D. G. (1994) Structure of Curacin A, a novel antimitotic, antiproliferative and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*, *The Journal of ...*
30. Chan, K.-S., Koh, C.-G., and Li, H.-Y. (2012) Mitosis-targeted anti-cancer therapies: where they stand, *Cell Death & Disease* 3, e411-.
31. Blokhin, A. V., Yoo, H. D., Gerald, R. S., Nagle, D. G., Gerwick, W. H., and Hamel, E. (1995) Characterization of the interaction of the marine cyanobacterial natural product curacin A with the colchicine site of tubulin and initial structure-activity studies with analogues., *Molecular Pharmacology* 48, 523-531.
32. Verdier-Pinard, P., Sitachitta, N., Rossi, J. V., Sackett, D. L., Gerwick, W. H., and Hamel, E. (1999) Biosynthesis of Radiolabeled Curacin A and Its Rapid and Apparently Irreversible Binding to the Colchicine Site of Tubulin, *Archives of Biochemistry and Biophysics* 370, 51-58.
33. Verdier-Pinard, P., Lai, J.-Y., Yoo, H.-D., Yu, J., Marquez, B., Nagle, D. G., Nambu, M., White, J. D., Falck, J. R., Gerwick, W. H., Day, B. W., and Hamel, E. (1998) Structure-Activity Analysis of the Interaction of Curacin A, the Potent Colchicine Site Antimitotic Agent, with Tubulin and Effects of Analogs on the Growth of MCF-7 Breast Cancer Cells, *molpharm.aspetjournals.org*.
34. Wipf, P., Reeves, J. T., Balachandran, R., and Day, B. W. (2002) Synthesis and Biological Evaluation of Structurally Highly Modified Analogues of the Antimitotic Natural Product Curacin A, *Journal of Medicinal Chemistry* 45, 1901-1917.
35. Wipf, P., Reeves, J. T., Balachandran, R., Giuliano, K. A., Hamel, E., and Day, B. W. (2000) Synthesis and Biological Evaluation of a Focused Mixture Library of Analogues of the Antimitotic Marine Natural Product Curacin A, *Journal of the American Chemical Society* 122, 9391-9395.
36. Gu, L., Wang, B., Kulkarni, A., Gehret, J. J., Lloyd, K. R., Gerwick, L., Gerwick, W. H., Wipf, P., Håkansson, K., Smith, J. L., and Sherman, D. H. (2009) Polyketide Decarboxylative Chain Termination Preceded by O-Sulfonation in Curacin A Biosynthesis, *Journal of the American Chemical Society* 131, 16033-16035.
37. Calderone, C. T. (2008) Isoprenoid-like alkylations in polyketide biosynthesis, *Natural product reports* 25, 845.
38. Gu, L., Eisman, E. B., Dutta, S., Franzmann, T. M., Walter, S., Gerwick, W. H., Skiniotis, G., and Sherman, D. H. (2011) Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect., *Angewandte Chemie (International ed. in English)* 50, 2795-2798.
39. Gu, L., Jia, J., Liu, H., Håkansson, K., Gerwick, W. H., and Sherman, D. H. (2006) Metabolic Coupling of Dehydration and Decarboxylation in the Curacin A Pathway: Functional Identification of a Mechanistically Diverse Enzyme Pair, *Journal of the American Chemical Society* 128, 9014-9015.
40. Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V., Gerwick, L., Håkansson, K., Wipf, P., Smith, J. L., Gerwick, W. H., and Sherman, D. H. (2009) Metamorphic enzyme assembly in polyketide diversification., *Nature* 459, 731-735.
41. Buchholz, T. J. (2010) Protein-Protein Interactions During Acyl Transfer In Polyketide Biosynthesis.

42. Pulsawat, N., Kitani, S., and Nihira, T. (2007) Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*., *Gene* 393, 31-42.
43. Piel, J., Wen, G., Platzer, M., and Hui, D. (2004) Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase., *ChemBioChem* 5, 93-98.
44. Calderone, C. T. (2006) Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*, *PNAS* 103, 8977-8982.
45. Jiang, H., Zirkle, R., Metz, J. G., Braun, L., Richter, L., Van Lanen, S. G., and Shen, B. (2008) The Role of Tandem Acyl Carrier Protein Domains in Polyunsaturated Fatty Acid Biosynthesis, *Journal of the American Chemical Society* 130, 6336-6337.
46. Crosby, J., and Crump, M. P. (2012) The structural role of the carrier protein–active controller or passive carrier, *Natural product reports* 29, 1111-1137.
47. Lebowitz, J., Lewis, M. S., and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review., *Protein Science* 11, 2067-2079.
48. Schuck, P. (2013) Analytical Ultracentrifugation as a Tool for Studying Protein Interactions., *Biophysical reviews* 5, 159-171.
49. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., and Weissman, K. J. (2003) The Structure of Docking Domains in Modular Polyketide Synthases, *Chemistry & Biology* 10, 723-731.
50. Buchholz, T. J., Geders, T. W., Bartley, F. E., Reynolds, K. A., Smith, J. L., and Sherman, D. H. (2009) Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains., *ACS Chemical Biology* 4, 41-52.
51. Hahn, M., and Stachelhaus, T. (2004) Selective interaction between nonribosomal peptidesynthetases is facilitated by shortcommunication-mediating domains, *Proceedings of the National Academy of Science* 101, 15585-15590.
52. Richter, C. D., Nietlispach, D., Broadhurst, R. W., and Weissman, K. J. (2007) Multienzyme docking in hybrid megasynthetases, *Nature Chemical Biology* 4, 75-81.
53. Busche, A., Gottstein, D., Hein, C., Ripin, N., Pader, I., Tufar, P., Eisman, E. B., Gu, L., Walsh, C. T., Sherman, D. H., Löhr, F., Güntert, P., and Dötsch, V. (2012) Characterization of Molecular Interactions between ACP and Halogenase Domains in the Curacin A Polyketide Synthase, *ACS Chemical Biology* 7, 378-386.
54. Ritchie, D. W. (2003) Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2, *Proteins: Structure, Function, and Bioinformatics* 52, 98-106.
55. Khare, D., Wang, B., Gu, L., Razelun, J., Sherman, D. H., Gerwick, W. H., Håkansson, K., and Smith, J. L. (2010) Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis., *PNAS* 107, 14099-14104.
56. Kapur, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2010) Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase., *PNAS* 107, 22066-22071.
57. Trujillo, U., Vázquez-Rosa, E., Oyola-Robles, D., Stagg, L. J., Vassallo, D. A., Vega, I. E., Arold, S. T., and Baerga-Ortiz, A. (2013) Solution structure of the tandem acyl carrier protein domains from a polyunsaturated fatty acid synthase reveals beads-on-a-string configuration., *PLoS ONE* 8, e57859.

58. Geders, T. W., Gu, L., Mowers, J. C., Liu, H., Gerwick, W. H., Hakansson, K., Sherman, D. H., and Smith, J. L. (2007) Crystal Structure of the ECH2 Catalytic Domain of CurF from *Lyngbya majuscula*: INSIGHTS INTO A DECARBOXYLASE INVOLVED IN POLYKETIDE CHAIN -BRANCHING, *Journal of Biological Chemistry* 282, 35954-35963.
59. Mendez-Perez, D., Begemann, M. B., and Pflieger, B. F. (2011) Modular Synthase-Encoding Gene Involved in α -Olefin Biosynthesis in *Synechococcus* sp. Strain PCC 7002, *Applied and Environmental Microbiology* 77, 4264-4267.
60. Gehret, J. J., Gu, L., Gerwick, W. H., Wipf, P., Sherman, D. H., and Smith, J. L. (2011) Terminal Alkene Formation by the Thioesterase of Curacin A Biosynthesis: STRUCTURE OF A DECARBOXYLATING THIOESTERASE, *Journal of Biological Chemistry* 286, 14445-14454.
61. Scaglione, J. B., Akey, D. L., Sullivan, R., Kittendorf, J. D., Rath, C. M., Kim, E.-S., Smith, J. L., and Sherman, D. H. (2010) Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase, *Angewandte Chemie International Edition* 49, 5726-5730.
62. Cramer, N. B., Davies, T., O'Brien, A. K., and Bowman, C. N. (2003) Mechanism and Modeling of a Thiol–Ene Photopolymerization, *Macromolecules* 36, 4631-4636.
63. Dondoni, A. (2008) The Emergence of Thiol-Ene Coupling as a Click Process for Materials and Bioorganic Chemistry, *Angewandte Chemie International Edition* 47, 8995-8997.
64. Hoyle, C. E., and Bowman, C. N. (2010) Thiol-Ene Click Chemistry, *Angewandte Chemie International Edition* 49, 1540-1573.
65. Sanchez, C., Du, L., Edwards, D. J., Toney, M. D., and Shen, B. (2001) Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide[^]polyketide antitumor drug bleomycin, *Chemistry & Biology*, 1-14.
66. McCarthy, J. G., Eisman, E. B., Kulkarni, S., Gerwick, L., Gerwick, W. H., Wipf, P., Sherman, D. H., and Smith, J. L. (2012) Structural basis of functional group activation by sulfotransferases in complex metabolic pathways., *ACS Chemical Biology* 7, 1994-2003.
67. Anderson, V. E., Bahnson, B. J., Wlassics, I. D., and Walsh, C. T. The reaction of acetyldithio-CoA, a readily enolized analog of acetyl-CoA with thiolase from *Zoogloea ramigera*., *jbc.org*.
68. Padmakumar, R., Gantla, S., and Banerjee, R. (1993) A rapid method for the synthesis of methylmalonyl-coenzyme A and other CoA-esters., *Analytical Biochemistry* 214, 318-320.
69. Arora, P., Goyal, A., Natarajan, V. T., Rajakumara, E., Verma, P., Gupta, R., Yousuf, M., Trivedi, O. A., Mohanty, D., Tyagi, A., Sankaranarayanan, R., and Gokhale, R. S. (2009) Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*, *Nature Chemical Biology* 5, 166-173.
70. Arora, P., Vats, A., Saxena, P., Mohanty, D., and Gokhale, R. S. (2005) Promiscuous Fatty Acyl CoA Ligases Produce Acyl-CoA and Acyl-SNAC Precursors for Polyketide Biosynthesis, *Journal of the American Chemical Society* 127, 9388-9389.
71. Rao, A., and Ranganathan, A. (2004) Interaction studies on proteins encoded by the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*, *Molecular Genetics and Genomics* 272, 571-579.

72. Léger, M., Gavaldà, S., Guillet, V., van der Rest, B., Slama, N., Montrozier, H., Mourey, L., Quémard, A., Daffé, M., and Marrakchi, H. (2009) The Dual Function of the Mycobacterium tuberculosis FadD32 Required for Mycolic Acid Biosynthesis, *Gene* 16, 510-519.
73. Zhang, Z., Zhou, R., Sauder, J. M., Tonge, P. J., Burley, S. K., and Swaminathan, S. (2011) Structural and functional studies of fatty acyl adenylate ligases from *E. coli* and *L. pneumophila*., *Journal of Molecular Biology* 406, 313-324.
74. Gavaldà, S., Léger, M., van der Rest, B., Stella, A., Bardou, F., Montrozier, H., Chalut, C., Burlet-Schiltz, O., Marrakchi, H., Daffé, M., and Quémard, A. The Pks13/FadD32 Crosstalk for the Biosynthesis of Mycolic Acids in Mycobacterium tuberculosis, *jbcb.org*.
75. Veyron-Churlet, R., Bigot, S., Guerrini, O., Verdoux, S., Malaga, W., Daffé, M., and Zerbib, D. (2005) The biosynthesis of mycolic acids in Mycobacterium tuberculosis relies on multiple specialized elongation complexes interconnected by specific protein-protein interactions., *Journal of Molecular Biology* 353, 847-858.
76. Mendez-Perez, D., Gunasekaran, S., Orlor, V. J., and Pfleger, B. F. (2012) A translation-coupling DNA cassette for monitoring protein translation in Escherichia coli, *Metabolic Engineering* 14, 298-305.
77. Burja, A. M., Banaigs, B., Abou-Mansour, E., Grant Burgess, J., and Wright, P. C. (2001) Marine cyanobacteria—a prolific source of natural products, *Tetrahedron* 57, 9347-9377.
78. Costa, M., Costa-Rodrigues, J., Fernandes, M. H., Barros, P., Vasconcelos, V., and Martins, R. (2012) Marine Cyanobacteria Compounds with Anticancer Properties: A Review on the Implication of Apoptosis, *Marine Drugs* 10, 2181-2207.
79. McPhail, K. L., Correa, J., Linington, R. G., González, J., Ortega-Barría, E., Capson, T. L., and Gerwick, W. H. (2007) Antimalarial Linear Lipopeptides from a Panamanian Strain of the Marine Cyanobacterium Lyngbyamajuscula, *Journal of Natural Products* 70, 984-988.
80. Liu, W., Ma, M., Xue, Y., Liu, N., Wang, S., and Chen, Y. (2013) The C-Terminal Extended Serine Residue Is Absolutely Required in Nosiheptide Maturation., *ChemBioChem*.
81. Yu, Y., Duan, L., Zhang, Q., Liao, R., Ding, Y., Pan, H., Wendt-Pienkowski, E., Tang, G., Shen, B., and Liu, W. (2009) Nosiheptide Biosynthesis Featuring a Unique Indole Side Ring Formation on the Characteristic Thiopeptide Framework, *ACS Chemical Biology* 4, 855-864.
82. Yu, Y., Guo, H., Zhang, Q., Duan, L., Ding, Y., Liao, R., Lei, C., Shen, B., and Liu, W. (2010) NosA Catalyzing Carboxyl-Terminal Amide Formation in Nosiheptide Maturation via an Enamine Dealkylation on the Serine-Extended Precursor Peptide, *Journal of the American Chemical Society* 132, 16324-16326.
83. Müller, I., and Müller, R. (2006) Biochemical characterization of MelJ and MelK., *FEBS Journal* 273, 3768-3778.
84. Müller, I., Weinig, S., Steinmetz, H., Kunze, B., Veluthoor, S., Mahmud, T., and Müller, R. (2006) A Unique Mechanism for Methyl Ester Formation via an Amide Intermediate Found in Myxobacteria, *ChemBioChem* 7, 1197-1205.
85. Jones, A. C., Monroe, E. A., Podell, S., Hess, W. R., Klages, S., Esquenazi, E., Niessen, S., Hoover, H., Rothmann, M., Lasken, R. S., Yates, J. R., Reinhardt, R., Kube, M., Burkart, M. D., Allen, E. E., Dorrestein, P. C., Gerwick, W. H., and Gerwick, L. (2011)

- Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*., *PNAS* 108, 8815-8820.
86. Dembitsky, V. M., Levitsky, D. O., Gloriovac, T. A., and Poroikov, V. V. (2006) Acetylenic Aquatic Anticancer Agents and Related Compounds, *Natural Product Communications* 1, 773-812.
 87. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases., *Chemistry & Biology* 6, 493-505.
 88. Apiradee, H., Kalyanee, P., Pongsathon, P., Patcharaporn, D., Matura, S., Sanjukta, S., Supapon, C., and Morakot, T. (2004) The expression of three desaturase genes of *Spirulina platensis* in *Escherichia coli* DH5alpha. Heterologous expression of *Spirulina*-desaturase genes., *Molecular Biology Reports* 31, 177-189.
 89. Sakamoto, T., Wada, H., Nishida, I., Ohmori, M., and Murata, N. (1994) delta 9 Acyl-lipid desaturases of cyanobacteria. Molecular cloning and substrate specificities in terms of fatty acids, sn-positions, and polar head groups., *The Journal of Biological Chemistry* 269, 25576-25580.
 90. Kurdrid, P., Subudhi, S., Hongsthong, A., Ruengjitchachawalya, M., and Tanticharoen, M. (2005) Functional Expression of *Spirulina*- Δ 6 Desaturase Gene in Yeast, *Saccharomyces cerevisiae*, *Molecular Biology Reports* 32, 215-226.
 91. Buist, P. H. (2004) Catalytic diversity of fatty acid desaturases, *Tetrahedron: Asymmetry* 15, 2779-2785.
 92. Higashi, S., and Murata, N. (1993) An in Vivo Study of Substrate Specificities of Acyl-Lipid Desaturases and Acyltransferases in Lipid Synthesis in *Synechocystis* PCC6803., *Plant Physiology* 102, 1275-1278.
 93. Hoyer, K. M., Mahlert, C., and Marahiel, M. A. (2007) The Iterative Gramicidin S Thioesterase Catalyzes Peptide Ligation and Cyclization, *Chemistry & Biology* 14, 13-22.
 94. Schmutz, E., Steffensky, M., Schmidt, J., Porzel, A., Li, S.-M., and Heide, L. (2003) An unusual amide synthetase (CouL) from the coumermycin A1 biosynthetic gene cluster from *Streptomyces rishiriensis* DSM 40489, *European Journal of Biochemistry* 270, 4413-4419.
 95. Steffensky, M. (2000) Cloning, Overexpression, and Purification of Novobiocin Acid Synthetase from *Streptomyces spheroides* NCIMB 11891, *Journal of Biological Chemistry* 275, 21754-21760.
 96. Das, D., Eser, B. E., Han, J., Sciore, A., and Marsh, E. N. G. (2011) Oxygen-Independent Decarbonylation of Aldehydes by Cyanobacterial Aldehyde Decarbonylase: A New Reaction of Diiron Enzymes, *Angewandte Chemie International Edition* 50, 7148-7152.
 97. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases., *Biochemistry* 37, 1585-1595.
 98. Quadri, L. E. N., Keating, T. A., Patel, H. M., and Walsh, C. T. (1999) Assembly of the *Pseudomonas aeruginosa* Nonribosomal Peptide Siderophore Pyochelin: In Vitro Reconstitution of Aryl-4,2-bisthiazoline Synthetase Activity from PchD, PchE, and PchF †, *Biochemistry* 38, 14941-14954.
 99. Jenke-Kodama, H. (2005) Evolutionary Implications of Bacterial Polyketide Synthases, *Molecular Biology and Evolution* 22, 2027-2039.

100. Bonnett, S. A., Rath, C. M., Shareef, A.-R., Joels, J. R., Chemler, J. A., Håkansson, K., Reynolds, K., and Sherman, D. H. (2011) Acyl-CoA Subunit Selectivity in the Pikromycin Polyketide Synthase PikAIV: Steady-State Kinetics and Active-Site Occupancy Analysis by FTICR-MS, *Chemistry & Biology* 18, 1075-1081.
101. Campbell, C. D., and Vederas, J. C. (2010) Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes, *Biopolymers* 93, 755-763.
102. Ma, S. M., Li, J. W. H., Choi, J. W., Zhou, H., Lee, K. K. M., Moorthie, V. A., Xie, X., Kealey, J. T., Da Silva, N. A., Vederas, J. C., and Tang, Y. (2009) Complete Reconstitution of a Highly Reducing Iterative Polyketide Synthase, *Science* 326, 589-592.
103. Evans, S. E., Williams, C., Arthur, C. J., Burston, S. G., Simpson, T. J., Crosby, J., and Crump, M. P. (2008) An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein, *ChemBioChem* 9, 2424-2432.
104. Płoskoń, E., Arthur, C. J., Kanari, A. L. P., Wattana-amorn, P., Williams, C., Crosby, J., Simpson, T. J., Willis, C. L., and Crump, M. P. (2010) Recognition of Intermediate Functionality by Acyl Carrier Protein over a Complete Cycle of Fatty Acid Biosynthesis, *Chemistry & Biology* 17, 776-785.
105. Ploskon, E., Arthur, C. J., Evans, S. E., Williams, C., Crosby, J., Simpson, T. J., and Crump, M. P. (2007) A Mammalian Type I Fatty Acid Synthase Acyl Carrier Protein Domain Does Not Sequester Acyl Chains, *Journal of Biological Chemistry* 283, 518-528.
106. Strieker, M., TanoviÄ, A., and Marahiel, M. A. (2010) Nonribosomal peptide synthetases: structures and dynamics, *Current Opinion in Structural Biology* 20, 234-240.
107. Becer, C. R., Hoogenboom, R., and Schubert, U. S. (2009) Click Chemistry beyond Metal-Catalyzed Cycloaddition, *Angewandte Chemie International Edition* 48, 4900-4908.
108. Sumerlin, B. S., and Vogt, A. P. (2010) Macromolecular Engineering through Click Chemistry and Other Efficient Transformations, *Macromolecules* 43, 1-13.
109. Barglow, K. T., and Cravatt, B. F. (2007) Activity-based protein profiling for the functional annotation of enzymes, *Nature Publishing Group* 4, 822-827.
110. Cravatt, B. F., Wright, A. T., and Kozarich, J. W. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry, dx.doi.org.
111. Jessani, N., and Cravatt, B. F. (2004) The development and application of methods for activity-based protein profiling, *Current Opinion in Chemical Biology* 8, 54-59.
112. Meier, J. L., Niessen, S., Hoover, H. S., Foley, T. L., Cravatt, B. F., and Burkart, M. D. (2009) An Orthogonal Active Site Identification System (OASIS) for Proteomic Profiling of Natural Product Biosynthesis, *ACS Chemical Biology* 4, 948-957.
113. Sieber, S. A., and Cravatt, B. F. (2006) Analytical platforms for activity-based protein profiling--exploiting the versatility of chemistry for functional proteomics., *Chemical communications (Cambridge, England)*, 2311-2319.
114. Speers, A. E., and Cravatt, B. F. (2003) Chemical Strategies for Activity-Based Proteomics, *ChemBioChem* 5, 41-47.
115. Hur, G. H., Vickery, C. R., and Burkart, M. D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology, *Natural product reports* 29, 1074.
116. Meier, J. L., Patel, A. D., Niessen, S., Meehan, M., Kersten, R., Yang, J. Y., Rothmann, M., Cravatt, B. F., Dorrestein, P. C., Burkart, M. D., and Bafna, V. (2011) Practical 4' -

- Phosphopantetheine Active Site Discovery from Proteomic Samples, *Journal of Proteome Research* 10, 320-329.
117. Kostiuk, M. A., Martin, D. D. O., Perinpanayagan, M. A., Hak, P. G., Siddam, A., Majjigapu, J. R., Rajaiah, G., Keller, B. O., Prescher, J. A., Wu, P., Bertozzi, C. R., Falck, J. R., and Berthiaume, L. G. (2009) Rapid and selective detection of fatty acylated proteins using ω -alkynyl-fatty acids and click chemistry, *J Lipid Res*, 1-44.
 118. Sidebottom, A. M., Johnson, A. R., Karty, J. A., Trader, D. J., and Carlson, E. E. (2013) Integrated Metabolomics Approach Facilitates Discovery of an Unpredicted Natural Product Suite from *Streptomyces coelicolor*M145, *ACS Chemical Biology* 8, 2009-2016.
 119. Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., Glukhov, E., Wodtke, A., de Felicio, R., Fenner, A., Wong, W. R., Linington, R. G., Zhang, L., Debonsi, H. M., Gerwick, W. H., and Dorrestein, P. C. (2013) Molecular Networking as a Dereplication Strategy, *Journal of Natural Products*, 130911151232007.