# BIOSYNTHETIC INNOVATION ON A POLYKETIDE ASSEMBLY LINE: BIOCHEMICAL STUDIES OF CURACIN A PATHWAY

**by**

**Liangcai Gu**

**A dissertation submitted in partial fulfillment**
**of the requirements for the degree of**
**Doctor of Philosophy**
**(Medicinal Chemistry)**
**in The University of Michigan**
**2008**

**Doctoral Committee:**

**Professor David H. Sherman, Chair**
**Associate Professor George A. Garcia**
**Professor Janet L. Smith**
**Professor Ronald W. Woodard**
**Assistant Professor Sylvie Garneau-Tsodikova**


CREDIT: R. CAMERON COATES

*To Li, and our parents*

# ACKNOWLEDGEMENTS

The first and foremost person I would like to thank is my advisor, Professor David Sherman. I deeply appreciate the freedom David gave me, and I also am grateful that he allowed me to work in the curacin project and pursue the topics I am interested in last four and a half years. I am especially thankful for the broad training I received from David. Without his support, encouragement, and guidance, I would never have had a productive doctoral research.

I would like to thank Professor Janet Smith, Professor Ronald Woodard, Professor George Garcia, and Professor Sylvie Garneau-Tsodikova for serving on my committee, and for all good suggestions on my projects.

I would like to thank Professor Janet Smith and Dr. Todd Geders in the Life Sciences Institute, Professor Kristina Håkansson, Bo Wang, and Dr. Haichuan Liu in Department of Chemistry for an exciting collaboration and stimulating discussions. I thank Professor Peter Wipf and Dr. Amol Kulkarni at University of Pittsburgh for synthesis of the substrates for my assays, Professor William Gerwick, Dr. Lena Gerwick, and Rashel Grindberg at University of California at San Diego for sending us the DNA of Jam enzymes for comparative studies of Cur and Jam systems, and Professor Christopher Walsh, and Dr. Christopher Calderone at Harvard Medical School for helpful discussions and sending us the $ACP_3$ constructs.

I greatly appreciate the immense help I received from the members in the Sherman group. I owe special thanks to Dr. Junyong Jia, who introduced me into the curacin project and taught me the skills of molecular cloning. Other labmates, such as Dr. Sabine Grüschow, Dr. Jeffrey Kittendorf, Dr. Courtney Aldrich, Shengyi Li, Yousong Ding, and

Shamilya Williams, offered me much help on both my research and study.

I would like to thank everyone with whom I had the pleasure to work with at University of Michigan: Dr. Kate Noon, Kari Anderson, Bryan Yestrepsky, Dr. Sergei Chernyak, Dr. Jaeman Byun, Dr. Dheeraj Khare, Jennifer Gehret, Jamie Razelun, Dr. David Akey, Dr. Sha Huang, Dr. Supratim Datta, Dr. Calvin Elmore, Mike Li, and all the others I've failed to list here.

Funding for my research came from NIH. I am also grateful for the support from the following fellowships and awards —Eli Lilly graduate fellowship, Rackham Predoctoral Fellowship, National Award for Outstanding Chinese Ph.D. Student Abroad, ACS Robert M. Scarborough Graduate Award, and travel grants from Rackham Graduate School and ACS.

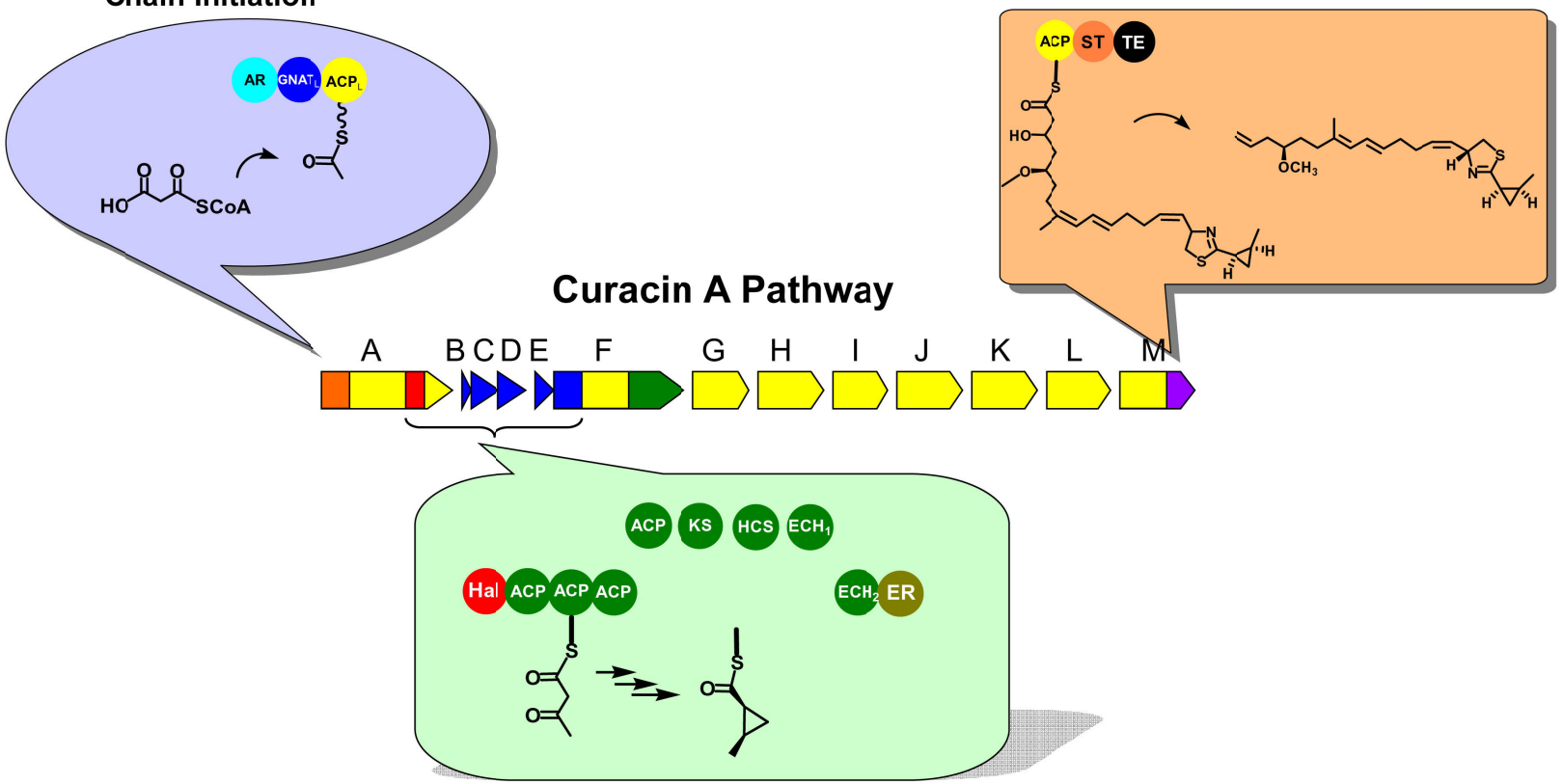
Last, but most importantly, I need to thank the support from my family members, especially Li, who has been the first audience of all my important presentations, who gives me whole-heart support for my doctoral study, and with whom our life and work in USA are meaningful.

# PREFACE

This dissertation contains six chapters regarding the biochemical studies of the chain initiation, HMG β-branching, halogenation, cyclopropanation and decarboxylative chain termination in the curacin A pathway (see next page for the structure of this dissertation). Chapter 1 is an introduction to the whole project, and some sections are partially adapted from our NIH proposal and an ecology manuscript submitted to *Applied and Environmental Microbiology.* Chapter 2 is mainly adapted from our paper published in Science regarding the chain initiation steps (*Science* **2007**, 318, 970-974). Chapter 3 is adapted from our two papers focused on the polyketide β-branching modification (*Journal of the American Chemical Society*, **2006**, 128, 9014-9015; *Journal of Biological Chemistry*, **2007**, 282, 35954-35963). Chapter 4 is adapted from our manuscript submitted to *Nature*, which is currently in revision. It tells a story about the cyclopropane formation and the vinyl chloride formation in the curacin and jamaicamide pathways, respectively. Chapter 5 describes a part of the work on the unusual chain termination, and this work is in preparation for publication. In chapter 6, I summarize some points we plan to discuss in a review article. Due to the limitation of space, results from some other projects I embarked on are not presented here.

Chapter 2: GNAT-Like Srategy for Polyketide Chain Initiation

Chapter 5: Decarboxylative Chain Termination




Chapter 3: Polyketide HMG β-Branching

Chapter 4: Halogenation, Cyclopropanation and Polyketide Diversification

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

## 1.1 BACKGROUND

### 1.1.1 Natural Products from Marine Cyanobacteria

Natural products have been playing a crucial role in the development of all classes of new pharmaceuticals, especially in the area of oncology. For example, of all cancer drugs available during the period 1940-2002, 40% are natural products or natural product derived, 10% are biologics or vaccines, 20% are synthetics that mimic natural product pharmacophores, and just 30% are of unique synthetic origin (*1*). Representatives for the National Cancer Institute's Developmental Therapeutics Program "strongly advocate expanding, not decreasing, the exploration of nature as a source of novel active agents that may serve as the leads and scaffolds for elaboration into desperately needed efficacious drugs for a multitude of disease indications."

Marine cyanobacteria are extraordinarily rich in their production of biologically-active and structurally-unique natural products, directly reflecting the amazing biosynthetic capacities of these organisms (*2, 3*). A number of these secondary metabolites represent lead compounds in drug discovery programs aimed at providing new therapies to treat cancer, bacterial infections, inflammatory responses and in crop protection to kill harmful microbial pathogens and insects. The major theme in cyanobacterial chemistry is the production of polypeptides modified with various lipid components to make diverse lipopeptides. Unusual functional groups are introduced into these lipopeptide scaffolds by action of numerous tailoring reactions, including those that form heterocyclic rings, oxidations at unusual positions, and halogenation reactions utilizing both bromine and chlorine, and occasionally, iodine.

The natural products isolated from both aquatic and marine cyanobacteria have been one of the richest sources of new clinical candidates and lead compounds for the treatment of cancer, and several of these exert their activity through modulation of microfilament polymerization processes. For example, a Curaçao collection of *Lyngbya majuscula* yielded a highly potent antitubulin agent, curacin A, which binds at the colchicine binding site on tubulin (*4, 5*). Curacin A has shown significant in vivo antitumor effects in human xenografts. However, significant problems with the instability and poor water solubility of curacin A have hindered its development as an anticancer lead compound (*6*). Wipf and his colleagues have been applying combinatorial synthesis methods to produce more stable and efficacious curacin A analogs (*6*).

### *1.1.2 Chemical Diversity of Lyngbya Natural Products*

The genus *Lyngbya* consists of filamentous cyanobacteria that cause periodic, but in some cases long lasting, blooms in shallow tropical and sub-tropical marine and estuarine environments (*7, 8*). *Lyngbya* species are prolific producers of secondary metabolites, primarily lipopeptides, cyclic peptides, and depsipeptides. Currently, the largest number of marine cyanobacterial metabolites comes from diverse strains of a single species, *L. majuscula* (Figure 1-1), obtained at tropical locations worldwide that produce collectively over 200 different compounds. Dissection of the biosynthetic subunits (e.g. amino acids, sugars, fatty acids) that comprise these metabolites indicates a range of 1 to 5 units per molecule, with about 25% of the precursors of these compounds derived from polyketides or fatty acids, another 25% from amino acids, and 25% from methyl groups likely deriving from S-adenosylmethionine (SAM). The remaining 25% of the biosynthetic subunits derive from terpenes and sugars, or are of uncertain origin. The amino acids are predominately aliphatic (70%), of L stereochemistry, *N*-methylated, and very rarely involve charged side chain residues. For example, many

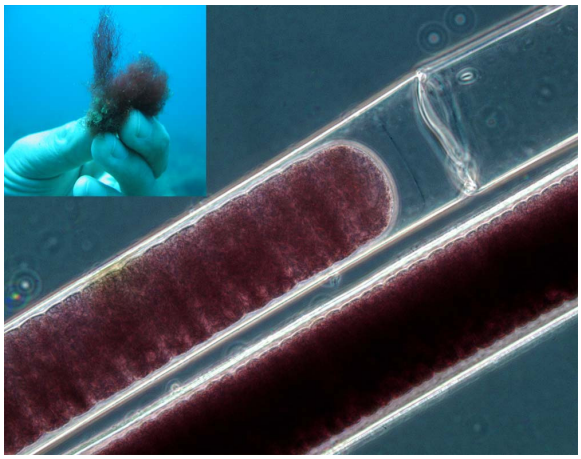


**Figure 1-1.** ***Lyngbya majuscula.***

marine cyanobacterial metabolites contain the unique "signature" amino acid, *N, O*-dimethyltyrosine. It can thus be concluded that the roughly 200 metabolites currently known from *L. majuscula* strains form an extraordinarily diverse set of natural product structures (*3*). As such, *L. majuscula* provides us an intriguing model system for genetic and biochemical studies towards understanding the evolution of the biosynthetic genes as well as metabolic diversification of natural products.

Many *Lyngbya*-derived bioactive secondary metabolites confer competitive advantage to the cyanobacteria by deterring grazers, allowing the cyanobacteria to overgrow other organisms in benthic habitats (*9*). In addition to rendering *Lyngbya* spp. unpalatable, bioactive peptides in *L.* spp. may have additional impacts, such as allelopathy against sympatric benthic organisms (*9, 10*). Also several bioactive compounds isolated from *L.* spp. have been well studied for their pharmaceutical and biotechnological applications, but very little is known about their ecological impacts on marine environments. Lyngbyatoxin A, an indole alkaloid first isolated from a Hawaiian strain of *L. majuscula* (*11*), is one of a few compounds subjected to extensive investigation for their ecological roles. It was proved to be a protein kinase C activator that causes contact dermatitis in humans and acts as a tumor promoter in mice (*12, 13*). In the marine habitat, it acts as a feeding deterrent to generalist grazers, but it stimulates feeding in the specialist grazer *Stylocheilus striatus* (*14*). Lyngbyatoxin A may have detrimental health impacts for generalist marine grazers that come into contact with the *Lyngbya* (*15*), and as a tumor promoter, it may be involved in fibropapillomatosis, a disease of marine turtles that causes internal and external tumorous masses (*16*). Thus, a better understanding of chemical diversity of *Lyngbya* has many implications for both ecosystem and human health.

### *1.1.3 Cyanobacterial PKS and NRPS Gene Clusters*

Marine cyanobacteria frequently combine polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) pathways to create an amazingly diverse set of natural product structures, generally characterized as "lipopeptides"(*2, 3, 17*). The biosynthetic gene clusters recently isolated from marine cyanobacteria by the Sherman and

Gerwick labs confirm the highly integrated use of PKS and NRPS pathways, and provide information about the unique biosynthetic machinery used by these microbes to introduce branching methyl groups, modified amino acids, and halogen atoms to produce diverse functionalities. To date, the Sherman and Gerwick labs have been involved in the complete cloning, sequencing and biochemical characterization of seven cyanobacterial pathways; barbamide (*bar*) (*18*) curacin A *(cur*) (*19*) cryptophycin (*crp*) (*20*), carmabin (*car*) (Jia *et al.*, unpublished), lyngbyatoxin (*ltx*) (*21*) jamaicamide (*jam*) (*22*), and hectochlorin (*hec*) (*23*) (Figure 1-2).

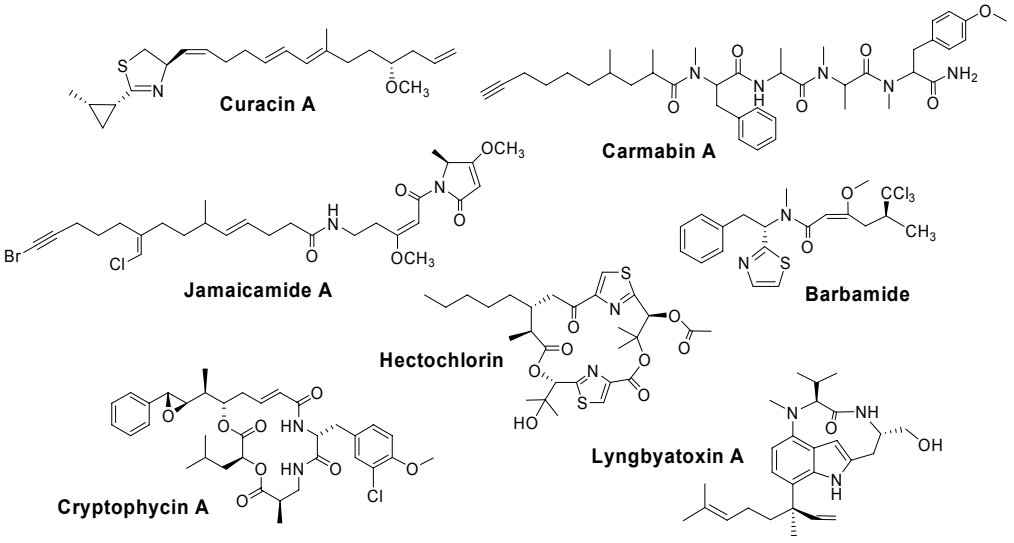



**Figure 1-2. Cyanobacterial metabolites with their biosynthetic pathways identified by the Sherman and Gerwick labs.**

With each new pathway, additional insights were developed into the unique nature of the secondary metabolite pathways of these organisms. They contain a wealth of novel gene sequences, gene cluster architectures and chain termination motifs. For example, in the curacin A and jamaicamide pathways (*19, 22*), a novel gene cassette was identified to be responsible for introduction of a branching carbon to a β-carbonyl compound (e.g. a diketide or a ketide-extended amino acid). According to protein database analysis, this eight open reading frame (ORF) cassette spanning ~10 kb of DNA is most closely related to an HMG-CoA synthase, enoyl-CoA hydratase along with several PKS-associated proteins, including tandem ACPs, and an unusual ketoacyl synthase (*24, 25*). Thus, consistent with

the feeding experiments, this novel gene arrangement condenses C-2 of an acetate unit with the β-carbonyl of a nascent polyketide, and by steps that remain largely undefined at present, catalyze the decarboxylation of C-1 resulting in the formation of a cyclopropyl ring in the case of curacin A,(*19*) and a vinyl chloride in the case of jamaicamide (*22*). Accordingly, literature review of known marine natural products provides evidence for the role of this gene cassette in the construction of over 50 cyanobacterial metabolites, introducing a wide range of functionalities in a pendant carbon atom (cyclopropyl, carbinol, secondary methyl, hydroxy methyl, vinyl methyl, terminal vinyl, and terminal vinyl chloride).

## *1.2 PRIOR WORK*

Curacin A is a unique polyketide/non-ribosomal peptide natural product discovered in 1994 (*26*). This marine cyanobacterial natural product has attractive features as an anticancer lead candidate and was shown (*27*) to have potent anti-proliferative and cytotoxic activity against colon, renal, and breast-cancer derived cell lines. Curacin A has been studied extensively for its ability to inhibit tubulin polymerization through interaction at the colchicine binding site, and thereby exert its antiproliferative effects on cells (*4*). Simultaneously, there has been considerable interest to produce curacin A by total synthesis (*28*), and to produce synthetic as well as semi-synthetic analogs to explore structure-activity relationships (*27*). Most recently, an elegant combinatorial synthesis approach has been applied that has resulted in a chemically-stabilized curacin A derivative with improved water solubility characteristics, yet it retains the remarkable biological effect of curacin A on microtubules (*6*).

Analysis of the precise biosynthetic origin of curacin A has been of great interest, and has provided key information to enable a complete genetic analysis of the curacin metabolic system (*19*). Precursor labeling studies using cultured *Lyngbya majuscula* demonstrated that the biosynthetic units for curacin A are acetate and cysteine (*29*). Notable structural features include a cyclopropane ring, thiazoline unit, *cis*-double bond and a terminal alkene residue. Each functional group is introduced by an intriguing set of biochemical steps, a subset of which we plan to study and then develop for chemoenzymatic synthesis of new metabolites.

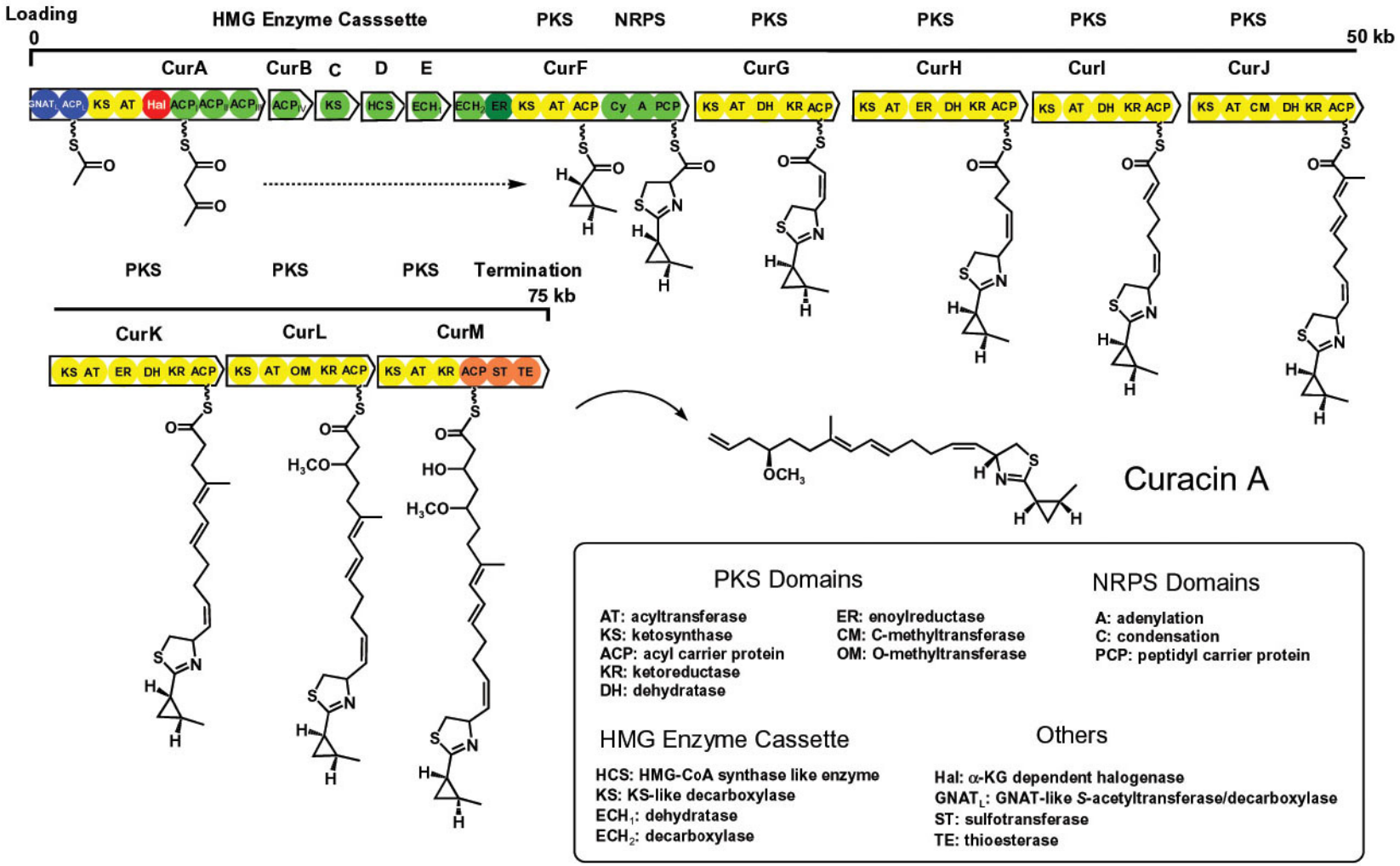



**Figure 1-3. Curacin A biosynthetic pathway**

Recently, we identified and characterized a gene cluster (*cur*) whose architecture and domain organization correlates directly with that expected for curacin A (*19*). A compelling feature of the system is its co-linear set of catalytic domains including a loading domain, a series of 14 ORFs (spanning ~75 kb) whose deduced sequences are consistent with biochemical elaboration of a cyclopropane ring, a single mixed PKS/NRPS bimodular unit, and a series of seven PKS monomodules with a terminal thioesterase (TE) domain (Figure 1-3). The *cur* cluster is unique in that each of the PKS multifunctional proteins is monomodular, a feature found in the final two modules of the pikromycin PKS (*30*). We plan to focus on several unique features of the curacin pathway. First, we are interested in several early steps including the sequence motif (Figure 1-3, Cur A (Hal-$ACP_I$-$ACP_{II}$-$ACP_{III}$), CurB, C, D, E, CurF (ECH-ER)) that likely prescribes synthesis of the cyclopropane ring and includes a stand-alone ACP, a putative HMG-CoA synthase-like enzyme (CurD), a ketosynthase and two enoyl-CoA hydratases (CurE $ECH_1$, CurF $ECH_2$ domain). The ability to understand the mechanistic basis for C-2 insertion reactions that provide variant functional group placements in natural product systems will be an important component of our planned biochemical studies of the jamaicamide system as well due to its identical architecture and extremely high level of amino acid sequence identity throughout this motif. Secondly, we are interested in understanding the chain initiation process which is most likely mediated by a domain showing homology to GNAT superfamily. Finally, we will propose experiments that focus on determining the basis for terminal double bond formation and release of curacin A, specified by the terminal catalytic domains of CurM.

## *1.3 SPECIFIC AIMS*

**1. To investigate the unusual polyketide chain initiation in the curacin A pathway.** By collaboration with the Smith group in Life Sciences Institute, we will try to understand the catalytic mechanisms of the corresponding enzymes.

**2. To investigate the cyclopropane formation in the curacin A pathway and the vinyl chloride formation in the jamaicamide pathway.** We need to establish the biochemistry of polyketide β-branching, halogenation and cyclopropanation. By

collaboration with the Smith group in Life Sciences Institute, we will try to understand the catalytic mechanisms of the corresponding enzymes.

**3. To investigate the decarboxylative chain termination in the curacin A pathway.** By collaboration with the Wipf group at the University of Pittsburgh, we will synthesize the substrates for the enzyme assays. By collaboration with the Smith group in Life Sciences Institute, we will try to understand the catalytic mechanisms of the corresponding enzymes.

## *1.4. References*


1. D. J. Newman, G. M. Cragg, K. M. Snader, *J. Nat. Prod*. **66**, 1022 (2003).
2. A. M. Burja, B. Banaigs, e. Abou-Mansour, J. G. Burgess, P. C. Wright, *Tetrahedron* **57**, 9347 (2001).
3. W. H. Gerwick, L. T. Tan, N. Sitachitta, *Alkaloids Chem. Biol.* **57**, 75 (2001).
4. A. V. Blokhin *et al.*, *Mol. Pharmacol*. **48**, 523 (1995).
5. E. Hamel, A. V. Blokhin, D. G. Nagle, H. Y. Yoo, W. H. Gerwick, *Drug Develop. Res*. **34**, 110 (1995).
6. P. Wipf, J. T. Reeves, R. Balachandran, B. W. Day, *J. Med. Chem*. **45**, 1901 (2002).
7. S. Albert *et al.*, *Mar. Pollut. Bull.* **51**, 428 (2005).
8. N. J. Osborne, G. R. Shaw, P. M. Webb, *Environ. Int.* **33**, 309 (2007).
9. V. J. Paul, K. E. Arthur, R. Ritson-Williams, C. Ross, K. Sharp, *Biol. Bull.* **213**, 226 (2007).
10. I. B. Kuffner *et al.*, *Mar. Ecol.-Progr. Ser.* **323**, 107 (2006).
11. J. H. Cardellina, F. J. Marner, R. E. Moore, *Science* **204**, 193 (1979).
12. H. Fujiki *et al.*, *Proc. Natl. Acad. Sci. USA*. **78**, 3872 (1981).
13. H. Fujiki *et al.*, *J. Cancer Res. Clin. Oncol.* **108**, 174 (1984).
14. A. Capper, I. R. Tibbetts, J. M. O'Neil, G. R. Shaw, *J. Exp. Ma. Biol. Ecol.* **331**, 133 (2006).
15. K. E. Arthur, C. J. Limpus, J. M. Whittier, *Aust. J. Zool.* **56**, 23 (2008).
16. K. Arthur *et al.*, *Harm. Algae* **7**, 114 (2008).
17. A. M. Burja *et al.*, *J Microbiol Method*. (2002), vol. 48, pp. 207-219.
18. Z. X. Chang *et al.*, *Gene* **296**, 235 (2002).
19. Z. Chang *et al.*, *J. Nat. Prod.* (2004), vol. 67, pp. 1356-1367.
20. N. A. Magarvey *et al.*, *ACS Chem. Biol.* **1**, 766 (2006).

21. D. J. Edwards, W. H. Gerwick, *J. Am. Chem. Soc.* (2004), vol. 126, pp. 11432-11433.
22. D. J. Edwards *et al.*, *Chem Biol*. (2004), vol. 11, pp. 817-833.
23. A. V. Ramaswamy, C. M. Sorrels, W. H. Gerwick, *J. Nat. Prod*. **70**, 1977 (2007).
24. Y. Paitan, E. Orr, E. Z. Ron, E. Rosenberg, *Microbiology*. (1999), vol. 145, pp. 3059-3067.
25. A. K. El-Sayed *et al.*, *Chem. Biol*. **10**, 419 (2003).
26. W. H. Gerwick *et al.*, *J. Org. Chem*. **59**, 1243 (1994).
27. P. Verdier-Pinard *et al.*, *Mol Pharmacol*. **53**, 62 (1998).
28. P. Wipf, W. Xu, in *J. Org. Chem.* **61**, 6556 (1996).
29. P. Verdier-Pinard *et al.*, *Arch. Biochem. Biophys*. **370**, 51 (1999).
30. Y. Xue, L. Zhao, H.-W.Liu, D. H. Sherman, *Proc. Natl. Acad. Sci. USA*. **95**, 12111 (1998).
31. L. E. N. Quadri *et al.*, *Biochemistry* **37**, 1585 (1998).

# CHAPTER 2
# GNAT-LIKE STRATEGY FOR POLYKETIDE CHAIN INITIATION

## 2.1 SUMMARY

In this chapter, a new biochemical strategy for chain initiation is described for the loading module of the curacin A polyketide synthase, an anti-cancer lead derived from the marine cyanobacterium *Lyngbya majuscula*. A central GNAT domain bears bifunctional decarboxylase/S-acetyltransferase activity, both unprecedented for the GNAT superfamily. A CurA loading tridomain consisting of adaptor-GNAT-acyl carrier protein was assessed biochemically, revealing that $GNAT_L$ catalyzes decarboxylation of malonyl-CoA to acetyl-CoA, and direct S-acetyl transfer from acetyl-CoA to $ACP_L$. Moreover, the N-terminal AR domain was shown to facilitate acetyl group transfer. Crystal structures of $GNAT_L$ were solved at 1.95Å (ligand-free form) and 2.75Å (acyl-CoA complex), showing distinct substrate tunnels for acyl-CoA and holo-$ACP_L$ binding. Modeling and site directed mutagenesis experiments demonstrated that $His^{389}$ and $Thr^{355}$ at the convergence of the CoA and ACP tunnels participate in malonyl-CoA decarboxylation, but not in acetyl transfer. Decarboxylation precedes acetyl transfer, leading to acetyl-$ACP_L$ as the key curacin A starter unit.

## 2.2 INTRODUCTION

Modular polyketide synthases (PKSs) are large biosynthetic machines that assemble structurally diverse secondary metabolites with a broad spectrum of biological activities. Multifunctional enzymes catalyze programmed metabolic pathways to assemble short-chain acyl-CoA building blocks into complex polyketide products by one step of chain initiation followed by multiple steps of chain elongation and processing. Polyketide

chain elongation steps are catalyzed by extension modules minimally comprised of three essential domains, ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP), plus auxiliary processing domains (e.g. ketoreductase, dehydratase, enoyl reductase). As in fatty acid synthases (FASs), PKSs catalyze chain extension by a decarboxylative condensation reaction. Polyketide chain initiation is catalyzed by a loading module that, in all characterized PKSs, consists of domains homologous with domains of the minimal extension module. A loading acyltransferase domain ($AT_L$) uses an acyl-CoA substrate to load an adjacent ACP domain ($ACP_L$). Typically, the $AT_L$ substrate is an α-carboxylated acyl-CoA, like the substrate for chain extension, and acyl transfer is followed by decarboxylation by a "$KS^Q$" domain within the loading module (*1, 2*). Alternatively, in the well studied erythromycin PKS, a *trans*-acting methylmalonyl-CoA decarboxylase acts on α-carboxylated acyl-CoA prior to acyl transfer by $AT_L$ (*3, 4*).

Curacin A, a marine cyanobacterial metabolite from *Lyngbya majuscula*, is a mixed polyketide-non-ribosomal peptide natural product with potent anti-proliferative and cytotoxic activity against colon, renal, and breast-cancer derived cell lines (*5*). The recent identification and partial characterization of the biosynthetic pathway for curacin A (*6*) revealed an atypical loading module, in which only the ACP domain ($ACP_L$) resembled typical PKS domains. Similar loading modules also occur in the biosynthetic pathways for pederin (*7*), its structural analogues onnamide/theopederin (*8*), myxovirescin A (*9*), and rhizoxin (*10*). In these natural product gene clusters, the chain initiation modules are structured as: N-terminal ~180 amino acid region of unknown function (referred to hereafter as adapter domain (AR), not present in OnnB for onnamide), a variable-length linker region, a domain showing homology to GCN5-related N-acyltransferase (GNAT) (hereafter referred to as $GNAT_L$), and an acyl carrier protein ($ACP_L$) (Figure 2-1). The predicted function of all of these $GNAT_L$-containing modules was to load an acetyl group onto the PKS assembly line (*6-10*), but the mechanism of this process remained unclear.

GNAT is a superfamily of N-acetyltransferase enzymes that catalyze acyl transfer to a primary amine and function in diverse pathways in prokaryotes and eukaryotes including antibiotic resistance, gene regulation, and hormone synthesis (*11, 12*).

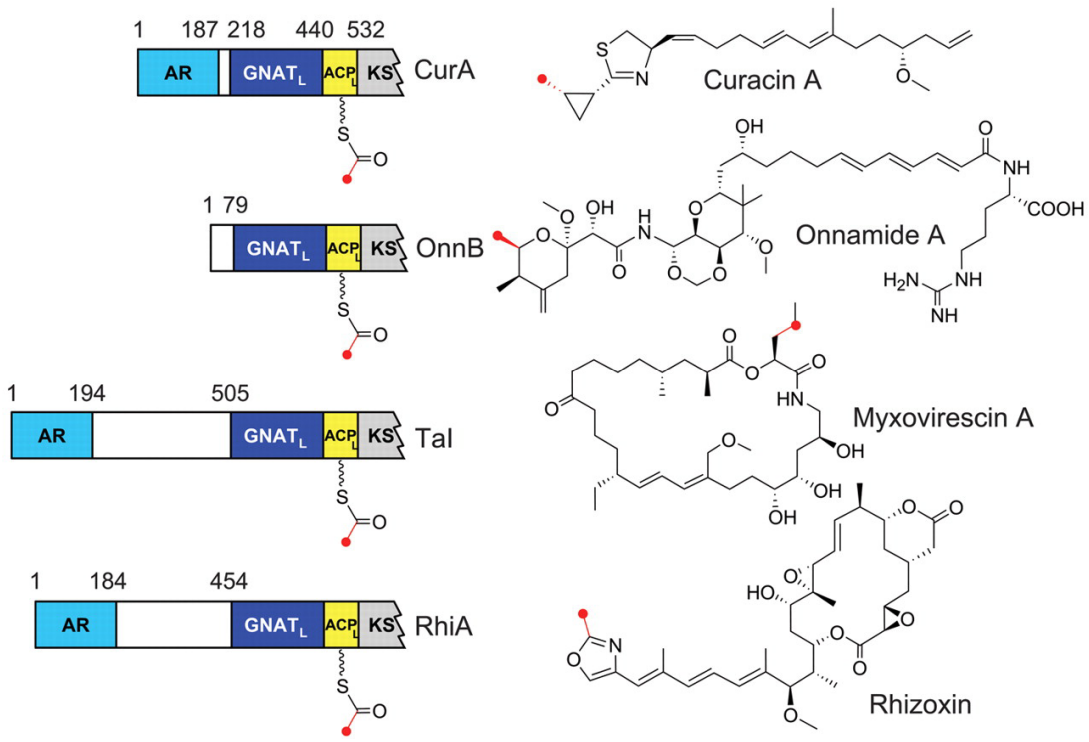



**Figure 2-1. Initiation modules containing GNAT$_L$ and compounds produced by the PKS pathways.** All GNAT$_L$ containing modules are predicted to catalyze the loading of an acetyl group. AR, adapter domain; GNAT$_L$, GNAT decarboxylase. Due to the high similarity of pederin, onnamides and theopederin, only onnamide A is shown as the prototype example.

Typically, prokaryotic and eukaryotic GNATs have separate binding sites for the acyl donor and acceptor substrates, and catalyze direct acyl transfer in the absence of a covalent intermediate on the enzyme (*11*). GNATs are mechanistically and structurally distinguished from the polyketide and fatty acid synthase acyltransferase domains that function as S-acyltransferases, employ a covalent enzyme intermediate, and belong to the α/β hydrolase superfamily (*13-16*). Thus, acyl transfer to an ACP thiol group represents an unprecedented reaction for GNAT enzymes.

## 2.3 RESULTS

### 2.3.1 Biochemical Assays of AR-GNAT$_L$-ACP$_L$ Tridomain

To investigate the atypical PKS chain initiation process, we cloned and overexpressed fragments of *curA* encoding the N-terminal tridomain, AR-GNAT$_L$-ACP$_L$.

Several expression constructs were made to include the three components in various combinations (Figure 2-2A and Figure 2-S1A, B). Constructs containing $ACP_L$ were

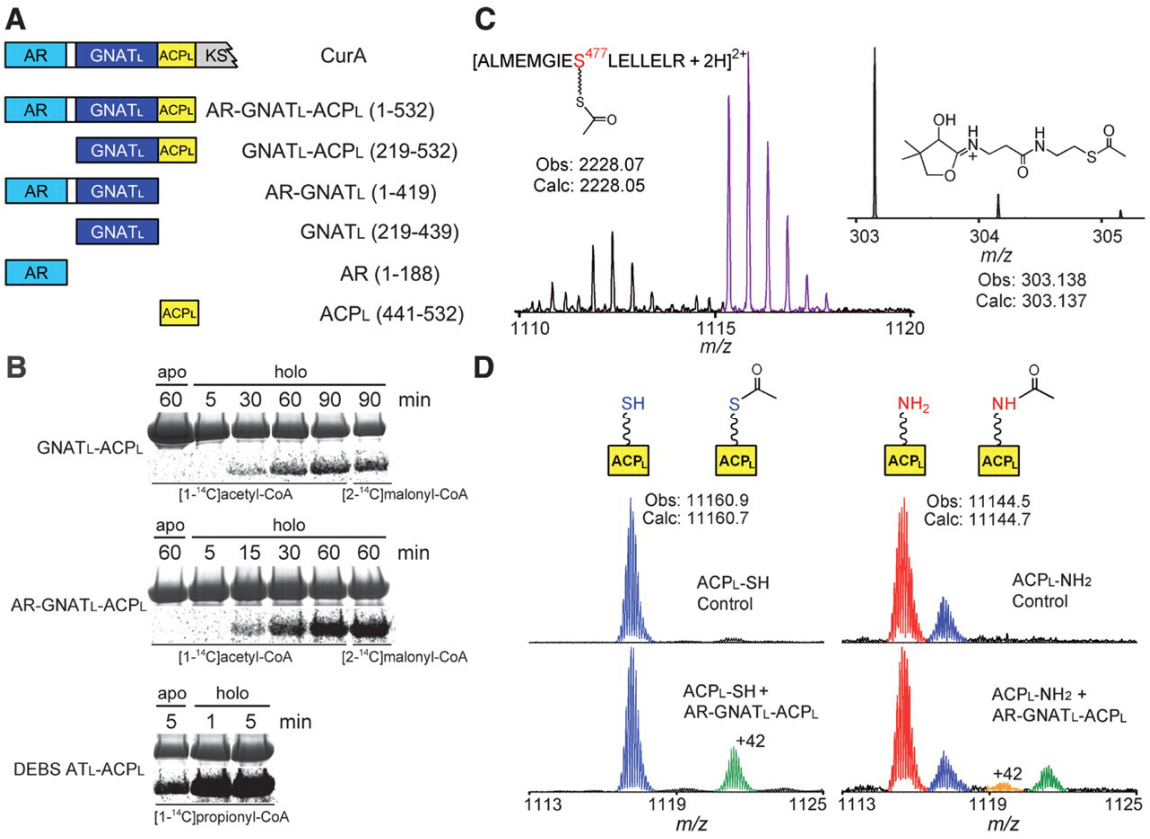



**Figure 2-2. Biochemical assays of *curA* initiation module.** (**A**) Expression constructs for *curA* initiation module. (**B**) Substrate loading of (apo and holo) $GNAT_L$-ACP, AR-$GNAT_L$-ACP and DEBS $AT_L$-$ACP_L$. Proteins (30 μM) were incubated with 90 μM CoA substrates, in 50 mM MOPS, pH 7.0, at room temperature. The SDS-PAGE gel images for Coomassie blue staining are shown above and autoradiography below. (**C**) FTICR mass spectrum (left) showing the ACP serine-containing peptide from the trypsin digested (holo) AR-$GNAT_L$-ACP loaded with malonyl-CoA, and partial IRMPD spectrum (right) showing the PPant ejection product with covalently-linked acetyl group. (**D**) Partial FTICR mass spectra showing (holo) $ACP_L$-SH (left) and (holo) $ACP_L$-$NH_2$ (right) loaded with acetyl-CoA. The N-terminal His-tag of $ACP_L$ was removed. 10 μM (holo) $ACP_L$-SH (or ~1.5 μM (holo) $ACP_L$-$NH_2$) and 2 μM AR-$GNAT_L$-$ACP_L$ were incubated with 50 μM acetyl-CoA at room temperature for 30 min.

generated in holo form by co-expression with a plasmid encoding phosphopantetheinyl transferase (Sfp) from *B. subtilis* (*17*); without this plasmid, *E. coli* produced $ACP_L$ in the apo form. All constructs lacking the AR domain were readily produced as soluble polypeptides, but those including AR (AR-$GNAT_L$ and AR-$GNAT_L$-$ACP_L$) had substantially decreased solubility, and the excised AR domain was not obtained in

soluble form under any conditions tested. Interestingly, solubility was significantly greater for the AR-$GNAT_L$-$ACP_L$ (holo) tridomain than for the apo counterpart (Figure 2-S1A), suggesting that the phosphopantetheine (PPant) arm of $ACP_L$ (holo) stabilizes the AR domain. The 63.3-kDa AR-$GNAT_L$-$ACP_L$ polypeptide is monomeric (Figure 2-S1C).

To assess the initiation behavior of the CurA starter unit, the apo and holo forms of the CurA tridomain (AR-$GNAT_L$-$ACP_L$) and didomain ($GNAT_L$-$ACP_L$) were treated with radio-labeled acyl-CoAs and analyzed by SDS-PAGE (Figure 2-2B). The holo forms of both the tridomain and didomain incorporated radiolabel, whereas the apo forms were not labeled, indicating transthioesterification of the acyl group from CoA to the PPant arm of ACP. Thus the N-terminal domains of CurA catalyze acyl loading and additional proteins are not required. Lack of radiolabeling of the apo forms also indicates that loading does not proceed through a covalent enzyme intermediate, consistent with the direct-transfer mechanism established for other GNAT family members (*3, 11*). This is in contrast to the "canonical" PKS loading module, typified by DEBS $AT_L$-$ACP_L$ (*3*), in which a covalent intermediate is observed for the apo form (Figure 2-2B). The ability of the $GNAT_L$-$ACP_L$ (holo) didomain to load an acyl group demonstrated that the catalytic machinery for chain initiation resides within the $GNAT_L$ domain and not the AR domain, which is also supported by the fact that selected mutations within $GNAT_L$ affected the loading behavior of AR-$GNAT_L$-$ACP_L$ both *in cis* (Figure 2-S2) and *in trans* (Figure 2-S6) (see below). However, the increased level of acyl-group loading activity by the AR-$GNAT_L$-$ACP_L$ (holo) tridomain (Figure 2-2B and Figure 2-S3) suggests that AR is required for efficient acyl transfer. Unexpectedly, both malonyl-CoA and acetyl-CoA functioned as substrates for the AR-$GNAT_L$-$ACP_L$ (holo) tridomain with similar efficiency (Figure 2-S2).

To investigate further the CurA AR-$GNAT_L$-$ACP_L$ tridomain components, we analyzed the products covalently tethered to the PPant arm of $ACP_L$. This was accomplished by interrogating mass changes on AR-$GNAT_L$-$ACP_L$ for *in cis* acyl transfer, or by employing the excised $ACP_L$ as the *in trans* acyl-group acceptor (Figure 2-2C). Excised $ACP_L$ and trypsin-digested AR-$GNAT_L$-$ACP_L$ samples were examined by

Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) and infrared multiphoton dissociation (IRMPD) methods (*18*). First, the $ACP_L$ phosphopantetheinylation site was established as $Ser^{477}$ by MS analysis of the trypsin-digested AR-$GNAT_L$-$ACP_L$ (Figure 2-S3), and by generating the corresponding S477A $ACP_L$ mutant protein. Remarkably, for both malonyl-CoA and acetyl-CoA substrates, only an acetyl group was detected on the PPant arm of AR-$GNAT_L$-$ACP_L$ (holo) (Figure 2-2C and Figure 2-S3) or $ACP_L$ (holo) (Figure 2-2D and Figure 2-S4A). Thus, the CurA $GNAT_L$ loading module catalyzes both decarboxylation and acyl transfer of carboxyl-acyl-CoA substrates. These data reveal a gain-of-function for a GNAT-type polypeptide, as well as a divergence from all other characterized PKS loading modules, in which the acyl transfer and decarboxylation are catalyzed by separate domains (*1, 4*). Further analysis by HPLC and FTICR-MS demonstrated that all constructs containing $GNAT_L$ ($GNAT_L$, $GNAT_L$-$ACP_L$ and AR-$GNAT_L$-$ACP_L$) catalyzed decarboxylation of malonyl-CoA, methylmalonyl-CoA, and malonyl-$ACP_L$ to generate acetyl-CoA, propionyl-CoA (Figure 2-S5) and acetyl-$ACP_L$ (Figure 2-S4B).

Next, kinetic parameters for decarboxylation were measured by HPLC for malonyl-CoA and methylmalonyl-CoA and by radio-assay for [1,3-$^{14}$C]malonyl-$ACP_L$ (Table 2-S1). The steady-state analysis indicated that malonyl-CoA is the preferred substrate. The $k_{cat}$ for malonyl-CoA was ~1.8 $s^{-1}$, which is ~6-fold and ~49-fold higher than those for methylmalonyl-CoA and malonyl-$ACP_L$, respectively. Similarly, the catalytic efficiency, $k_{cat}/K_M$, for malonyl-CoA was ~5.25 $mM^{-1}s^{-1}$, which is ~3-fold and ~6-fold higher than those of methylmalonyl-CoA and malonyl-$ACP_L$, respectively.

With clear evidence for a relatively rapid decarboxylation step catalyzed by $GNAT_L$, we sought to measure the rate of acyl loading by AR-$GNAT_L$-$ACP_L$. The *in cis* acetyl transfer rate of AR-$GNAT_L$-$ACP_L$ (holo) was determined by using [1-$^{14}$C]acetyl and [2-$^{14}$C]malonyl-CoA substrates. The $k_{cat}$ and $K_M$ values were derived by measuring the intramolecular acyl transfer rate at a series of acyl-CoA concentrations (Table 2-S1). The $k_{cat}$ or $K_M$ values for acetyl transfer were similar for acetyl-CoA and malonyl-CoA substrates. In contrast, the $k_{cat}$ for acetyl-group transfer was ~780-fold slower than the $k_{cat}$ for decarboxylation of malonyl-CoA, suggesting that decarboxylation and acetyl transfer

are separated by a slow tri-domain conformational change that leads to effective binding of the $ACP_L$ PPant arm in the active site ($k_2$ in Figure 2-S7). Taken together with the decarboxylation kinetic data, these results confirm that malonyl-CoA decarboxylation precedes acetyl transfer to $ACP_L$ (holo). In addition, the apparent $K_M$ of acetyl-CoA or malonyl-CoA for acyl transfer is 80-90 fold lower than the $K_M$ of malonyl-CoA for decarboxylation, and is dependent on the ratio of rate constants ($k_1$-$k_6$ in Figure 2-S7).

Since previously described GNAT enzymes catalyze various N-acetylation reactions, we tested whether CurA $GNAT_L$ retained a similar residual activity. We synthesized a CurA $ACP_L$-phosphopantetheine analog bearing a terminal $NH_2$ in place of native (holo) $ACP_L$-SH (refer to SOM), and tested it for *in trans* acetyl transfer catalyzed by AR-$GNAT_L$-$ACP_L$ (apo) to excised CurA (holo) $ACP_L$-$NH_2$ (monitored by FTICR-MS). A convenient internal standard was supplied by low levels of $ACP_L$-SH in the $ACP_L$-$NH_2$ preparation (Figure 2-2D), likely due to the activity of endogenous *E. coli* ACP synthase or EntD (*19*). In contrast to the significant level of acetyl loading on $ACP_L$-SH (Figure 2-2D), only a trace amount of acetyl-NH-$ACP_L$ was detected. In addition, simple alkyl amines (e.g. ethylenediamine, butylamine) were tested as substrates for acetyl transfer by CurA $GNAT_L$, but no N-acetyl transferase activity was detected for any of them. Thus, the *N*-acetyltransfer activity typically associated with GNAT enzymes is almost completely attenuated in the Cur AR-$GNAT_L$-$ACP_L$ chain initiation module.

### 2.3.2 $GNAT_L$ Structure and Site-Directed Mutagenesis

To advance our understanding of $GNAT_L$ function and to identify active site residues, we determined crystal structures of the excised CurA apo $GNAT_L$ (ligand free) domain and of the corresponding $GNAT_L$ with added malonyl-CoA (Table 2-S2). $GNAT_L$ (CurA residues 219-439) possesses the GNAT superfamily fold, consisting of a central mostly anti-parallel β-sheet flanked by α-helices (Figure 2-3A) (*11, 12*), and is most similar to serotonin N-acetyltransferase (RMSD = 1.9Å for 160 $C_\alpha$ atoms) (*20*).

The crystal structures of CurA $GNAT_L$ provide key insights into the function of this unique member of a large and ubiquitous protein family. Two tunnels from opposite faces of the protein converge at a position corresponding to the active sites of homologous

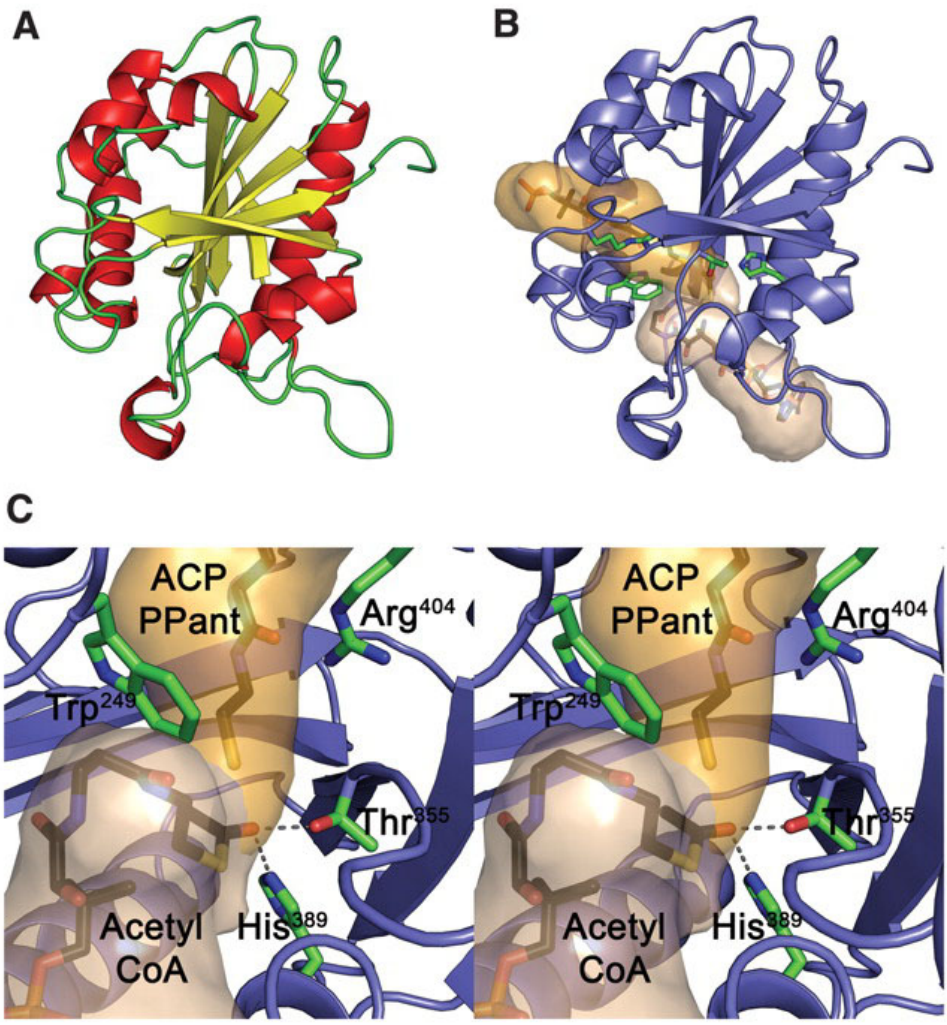



**Figure 2-3. Structure of $GNAT_L$ domain.** (**A**) Structural fold of $GNAT_L$ domain. (**B**) Substrate tunnels with models (black carbons) of acetyl-CoA (wheat) and PPant arm of holo-ACP (orange). Residues $Trp^{249}$, $His^{389}$, $Thr^{355}$, and $Arg^{404}$ are shown in green carbons. (**C**) Stereo diagram of modeling results overlaid with observed structure showing interaction of $His^{389}$ and $Thr^{355}$ with thioester carbonyl of acetyl-CoA. Coloring is identical to (**B**).

GNATs (Figure 2-3B). Presumably these tunnels are the binding sites for the phosphopantetheine (PPant) arms of the CoA and ACP substrates. To distinguish the two tunnels, crystals of $GNAT_L$ were soaked with malonyl-CoA (Table 2-S2). New electron density was observed in only one tunnel, hereafter called the CoA binding tunnel (Figure 2-S8). The location of the tunnel and the mode of CoA binding, in which the nucleotide lies in a surface cleft and the PPant arm extends into the tunnel, are consistent with structures of other members of the GNAT superfamily (*11, 12*). In CurA $GNAT_L$,

selectivity for CoA over ACP is imparted by a 5'-diphosphate binding loop (residues 327-332) and by $Arg^{387}$ recognition of the 3'-phosphate. Neither of these recognition features exists in FeeM, a GNAT superfamily member that uses an acyl-ACP donor substrate (*21*). Compared to the strong density for the CoA nucleotide, density for the PPant arm was weaker and indicative of multiple conformations, perhaps due to presence of both the malonyl-CoA substrate and the acetyl-CoA decarboxylation product in the crystal. No new electron density was observed in the second tunnel, which we designate the ACP-binding tunnel. The side chain of conserved $Arg^{404}$ forms one wall of the putative ACP-binding tunnel (Figure 2-3C and Figure 2-S9), and a series of water molecules extending to the protein exterior form the opposite wall. The designated ACP-binding tunnel overlays well with the acetyl group acceptor site in other GNATs, is well matched in length to the fully extended PPant arm of holo ACP (~15Å), and likely forms the $ACP_L$ PPant arm binding site (Figure 2-3B). The two tunnels meet at conserved $Trp^{249}$, just as many other GNAT family members have their two binding sites separated by an aromatic residue (Figure 2-3B and Figure 2-S9).

We searched for potential catalytic residues for decarboxylation and found $His^{389}$ and $Thr^{355}$ at the junction of the two tunnels (Figure 2-S9 and Figure 2-3C). These residues are precisely positioned by hydrogen bonds of $His^{389}$ to $Tyr^{419}$, and of the $Thr^{355}$ backbone to the $Arg^{404}$ side chain. All four of these residues are invariant among $GNAT_L$-containing PKS loading modules (Figure 2-S9). To test the role of $His^{389}$ and $Thr^{355}$, H389A, H389N and T355V variant proteins were assayed for decarboxylase and acetyltransferase activity (Table 2-S1). Decarboxylation was severely impaired by substitutions at either site. Specifically, the $His^{389}$ and $Thr^{355}$ variants resulted in at least 100-fold reduction in $k_{cat}$ for decarboxylation of malonyl-CoA to acetyl-CoA and in relatively modest changes to $K_M$ values. In contrast, substitutions at $His^{389}$ and $Thr^{355}$ had only modest effects on acetyl transfer activity, demonstrating that these residues do not play a critical role in acetyl transfer.

Based on mutagenesis results and on modeling the PPant arms of acetyl-CoA and $ACP_L$ (Figure 2-3C and Figure 2-S8), $His^{389}$ and $Thr^{355}$ are proposed to stabilize an enolate anion intermediate of the decarboxylation reaction. We further propose that the

isoenergetic transfer of an acetyl group from acetyl-CoA to $ACP_L$-SH (holo) is catalyzed by direct attack of the deprotonated thiolate of $ACP_L$ upon acetyl-CoA at the junction of the $GNAT_L$ CoA and $ACP_L$ binding tunnels. The buried $Arg^{404}$ residue at the base of the $ACP_L$ binding tunnel may facilitate binding of the deprotonated thiol of the acceptor PPant arm of $ACP_L$. The N-terminal AR domain, which was stabilized by the PPant arm of $ACP_L$, may also assist in delivering (holo) $ACP_L$ to $GNAT_L$. AR is likely to have a common function in $GNAT_L$-containing modules based on its highly conserved sequence.

Based on the studies described above, the mechanism of CurA AR-$GNAT_L$-$ACP_L$ is proposed to involve a series of acyl-CoA-protein, protein-protein, and protein-PPant arm interactions (Figure 2-4 and Figure 2-S7) that mediate the unique bifunctional decarboxylase/S-acetytransferase activity. It is likely that other novel GNAT members will be found with surprising biochemical properties or that previously identified proteins will be grouped within this large family. For example, the *eryM*-encoded methylmalonyl-CoA decarboxylase (*4*) has not been the subject of structural analysis, but comparative amino acid sequence analysis now predicts that it contains the GNAT scaffold.

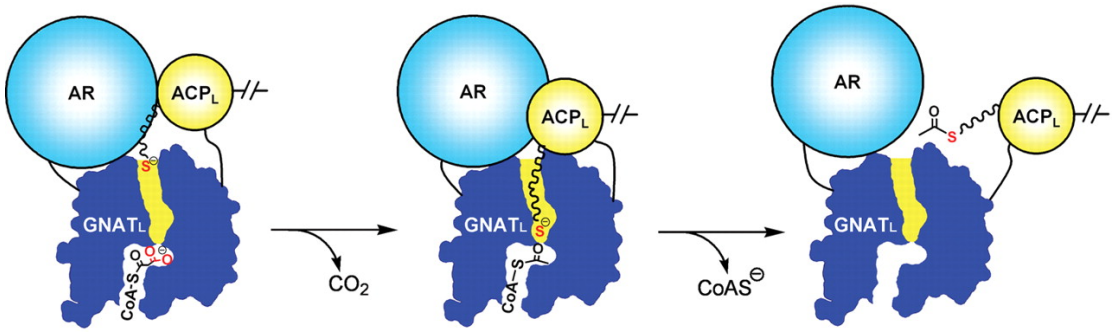



**Figure 2-4. Proposed mechanism of the CurA AR-$GNAT_L$-$ACP_L$ chain initiation module.** Malonyl-CoA enters the $GNAT_L$ domain (blue) CoA tunnel (white) and catalyzes decarboxylation to acetyl-CoA. The AR domain (light blue) directs the phosphopantetheine arm of $ACP_L$ (yellow) into the ACP tunnel (yellow) for subsequent acetyl group transfer via transthioesterification.

## 2.4 DISCUSSION

A significant outcome of the current work is the realization that $GNAT_L$-containing

modules are not uncommon among previously characterized natural product biosynthetic gene clusters, and are well represented within bacterial genome sequences. This metabolic strategy thus represents an additional widely employed chain initiation process for assembly of important biologically active small molecules. At a comparative level, the CurA $GNAT_L$ has a significantly slower S-acetyltransfer activity than the erythromycin PKS $AT_L$. However, *L. majuscula* has a relatively slow growth rate in the marine environment as well as in culture (i.e. doubling time of approximately 10 days) (*22*), and a correspondingly low production of curacin A per unit biomass compared to yields of other microbial natural products (e.g. erythromycin via the 6-deoxyerythronolide B aglycone intermediate) by *Saccharopolyspora erythraea* (*23*)). Thus, the curacin A PKS assembly line has evolved under environmental constraints and organismic needs that reflect its reduced efficiency.

The work described in this report provides clear evidence for an unprecedented bifunctional decarboxylase/S-acetyltransferase role for the CurA GNAT scaffold, which significantly broadens the chemical reaction inventory of this well known protein superfamily. This newly described method of PKS chain initiation enables malonyl CoA to serve as the sole precursor to both initiate and extend the carbon chain of curacin A. Moreover, by virtue of the greater rate of acetate versus propionate transfer to $ACP_L$ following decarboxylation, the $GNAT_L$ strategy for chain initiation provides an additional mechanism to insure fidelity of curacin A chain length.

## *2.5 MATERIAL AND METHODS*

**Chemicals.** [1,3-$^{14}$C]malonyl-CoA (55 mCi/mmol), [2-$^{14}$C]malonyl-CoA (55 mCi/mmol), and [1-$^{14}$C]acetyl-CoA (55 mCi/mmol) were from American Radiolabeled Chemicals. [1-$^{14}$C]propionyl-CoA (55 mCi/mmol) was from Moravek Biochemicals. The terminal-$NH_2$ analog of pantetheine was synthesize by Michael D. Burkart group at UCSD (*24*). All other chemicals were from Sigma-Aldrich.

**Bacterial strains, media and culture conditions.** *Escherichia coli* DH5α MCR (Invitrogen) or XL1-Blue (Stratagene) was used for DNA propagation. *Escherichia coli* BL21 (DE3) and Rosetta (DE3) (Invitrogen) transformed with the derivatives of pET24b,

pET28b, pET21b (Novagen) and pMCSG7 (*25*) were used for protein overexpression in Luria-Bertani (LB) medium. Ampicillin (100 μg/mL), carbenecillin (100 μg/mL), kanamycin (50 μg/mL), chloramphenicol (25 μg/mL), and apramycin (50 μg/mL) were used for the corresponding plasmid construct resistance marker selection in *E. coli* cultures.

**Plasmid construction and site mutagenesis.** The primers used for the plasmid construction and site mutagenesis are listed in table 2-S3. AR-$GNAT_L$-$ACP_L$, AR-$GNAT_L$, $GNAT_L$-$ACP_L$, $GNAT_L$, $ACP_L$ and AR genes were amplified from cosmid pLM54 described in our previous study (*6*). DEBS loading $AT_L$-$ACP_L$ gene was amplified from cosmid pBK3 previously described (*26*). AR-$GNAT_L$-$ACP_L$, $GNAT_L$-$ACP_L$, and $ACP_L$ genes were inserted into pET28b plasmid. AR-$GNAT_L$, $GNAT_L$ and AR genes were inserted into pET24b plasmid. DEBS loading $AT_L$-$ACP_L$ gene was inserted into pET21b plasmid similarly as reported (*3*). All of these pET derivatives were constructed with *Nde*I and *Xho*I restriction sites. H389A, H389N and T355V site mutagenesis for both $GNAT_L$ and AR-$GNAT_L$-$ACP_L$ was performed using QuikChange site mutagenesis protocol (Stratagene). All the constructs and mutations were verified by DNA sequencing.

**Protein overexpression. A) (apo or holo) AR-$GNAT_L$-$ACP_L$, $GNAT_L$-$ACP_L$, $ACP_L$** and **corresponding mutants:** *E. coli* BL21 (DE3) was transformed by the pET28b::AR-$GNAT_L$-$ACP_L$, pET28b::$GNAT_L$-$ACP_L$ or pET28b::$ACP_L$ plasmid to overexpress N-terminal His-tagged proteins. The proteins in holo-form were produced by coexpression with Sfp plasmid pSG701 (untagged Sfp, under T7 control, chloramphenicol resistance). Cells were grown at 30$^o$C to an OD (590 nm) = 0.5-0.6, and then cooled to 16$^o$C prior to the addition of 1 mM isopropyl-β-D-galactopyranoside (IPTG). The cultures were grown at 16$^o$C for another 18-20 h before harvesting. **B) AR-$GNAT_L$, $GNAT_L$**, and **corresponding mutants:** *E. coli* BL21 (DE3) was transformed by the pET24b::AR-$GNAT_L$ or pET24b::$GNAT_L$ plasmid to overexpress C-terminal His-tagged proteins. Cells were grown at 35$^o$C to an OD (590 nm) = 0.5-0.6, and then cooled to 18$^o$C prior to the addition of 1 mM IPTG. The cultures were grown at 18$^o$C for another 12-15 h before harvesting. **C) $GNAT_L$ for crystallization:** *E. coli* BL21(DE3) was transformed by the pET28b::$GNAT_L$ or pMCSG7::$GNAT_L$ plasmid to overexpress C- or N-terminal His-tagged proteins, respectively. Cells were grown at 37 $^o$C to an OD (600 nm) = 0.6-0.8,

and then cooled to 18°C prior to the addition of 0.4 mM IPTG. The cultures were grown at 18°C for another 12-15 h before harvesting. **D) (apo or holo) DEBS $AT_L$-$ACP_L$:** *E. coli* Rosetta (DE3) (chloramphenicol resistance) was transformed by pET21b::$AT_L$-$ACP_L$ (ampicillin resistance) to overexpress the C-terminal His-tagged proteins. (Holo) $AT_L$-$ACP_L$ was generated by coexpression with Sfp plasmid pRSG56 (kanamycin resistance) (*27*). Cells were grown at 37°C to an OD (590 nm) = 0.5-0.6, and then cooled to 30°C prior to the addition of 1 mM IPTG. The cultures were grown at 30°C for another 8-10 h before harvesting. **E) *E. coli* CoAA** and ***H. sapiens* PPAT–DPCK:** *E. coli* BL21 (DE3) was transformed by CoAA (kanamycin resistance) or PPAT–DPCK (ampicillin resistance) expression plasmid (gift of Prof. Michael D. Burkart, UCSD) to express His-tagged proteins (*28*). Cells were grown at 37°C to an OD (590 nm) = 0.7-0.8, and then cooled to 20°C prior to the addition of 0.8 mM IPTG. The cultures were grown at 20°C for another 12-15 h before harvesting.

**Protein purification.** Protein purifications were performed at 4°C. Generally, the first step Ni-affinity purifications for all His-tagged proteins were performed under the same conditions. *E. coli* cells were harvested by centrifugation (4,000 g, 15 min, 4°C), resuspended in the ice cold lysis buffer A (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol) and disrupted by sonication on ice. The cell debris was removed by centrifugation at 15,000 g for 50 min. The supernatant was gently removed and loaded onto the 5 mL HisTrap column (GE Healthcare) preequilibrated with lysis buffer A. The resin was washed successively with ~ 10 column volumes of the washing buffer B (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol) to remove nonspecifically bound contaminants. Bound proteins were eluted with imidazole by a linear gradient of the elution buffer C (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 250 mM imidazole, 20% glycerol). The eluate was concentrated using Amicon Ultra-15 (30 kDa, 10 kDa or 5 kDa) centrifugal devices (Millipore). The follow-up purification and buffer exchange for different proteins are described below:

**A) (apo or holo) AR-$GNAT_L$-$ACP_L$/$GNAT_L$-$ACP_L$, AR-$GNAT_L$/$GNAT_L$** and **the corresponding mutants:** the concentrated eluate was loaded onto HiLoad 26/60 Superdex 200 column (GE Healthcare) equilibrated with the storage buffer D (50 mM PBS buffer,

pH 7.5, 200 mM NaCl, 20% glycerol). The fractions were pooled, concentrated, flash-frozen in 50-100 μL aliquots in liquid $N_2$, and stored at -80°C for further use. **B) (apo or holo) DEBS $AT_L$-$ACP_L$:** the concentrated eluate was loaded onto HiPrep 26/10 Desalting column (GE Healthcare) equilibrated with the storage buffer D. The fractions were pooled, concentrated, flash-frozen in 50 μL aliquots in liquid $N_2$, and stored at -80°C for further use. **C) CoAA** and **PPAT–DPCK:** the concentrated eluate was loaded onto HiPrep 26/10 Desalting column equilibrated with the storage buffer E (50 mM HEPES, pH 7.5, 0.1 M NaCl, 0.5 mM EDTA, 1 mM DTT and 20% glycerol). The fractions were pooled, concentrated, flash-frozen in 50 μL aliquots in liquid $N_2$, and stored at -80°C for further use. **D) His-tag cleaved (apo or holo) $ACP_L$**: the N-terminal His-tag of $ACP_L$ needs to be removed to avoid a 178 Da posttranslational modification in *E. coli* which was reported to be the α-N-6-phosphogluconoylation on N-terminal His-tag (*29*). Firstly, the concentrated eluate was loaded onto HiPrep 26/10 Desalting column equilibrated with the storage buffer D. The biotinylated thrombin (Novagen) was added into the pooled fractions, and the reaction was incubated at room temperature for overnight to achieve a complete cleavage. The reaction mixture was loaded onto 1mL HiTrap Streptavidin HP (GE Healthcare) to trap the biotinylated thrombin and the column was successively washed with 10 mL storage buffer D. The flow-through fractions were pooled, concentrated and reloaded onto HiPrep 26/10 Desalting column equilibrated with the storage buffer E. The fractions were pooled, concentrated, flash-frozen in 100 μL aliquots in liquid N2, and stored at -80°C for further use.

The purity of the proteins was analyzed by SDS-PAGE and the protein concentrations were determined using the Bradford assay (Bio-Rad). The percentage of the holo form of $ACP_L$ was analyzed by reverse-phase HPLC using a Jupiter C4 column (250 x 2.0 mm, 5 μm, 300 Å, Phenomenex). The samples were eluted with a linear gradient from 5% to 90% of $CH_3CN$ (0.1% $CF_3CO_2H$)/$H_2O$ (0.1% $CF_3CO_2H$). According to the HPLC assay, 60-70% $ACP_L$ was converted into holo form by coexpression with Sfp. The holo/apo ratio of AR-$GNAT_L$-$ACP_L$ samples was analyzed by trypsin digestion followed by FTICR-MS analysis. More than 90 % AR-$GNAT_L$-$ACP_L$ was converted into holo form by coexpression with Sfp.

**Preparation of $ACP_L$-$NH_2$.** One-pot chemo-enzymatic strategy (*28*) was used to synthesize $ACP_L$-$NH_2$. To simplify the $ACP_L$-$NH_2$ purification, we did not use pyruvate kinase and inorganic pyrophosphatase to recycle the ATP. Briefly, a 200 μL reaction mixture contains: 20 μL 10x Tris buffer (500 mM, pH 8.0), 20 μL $MgCl_2$ (100 mM), 20 μL KCl (200 mM), 70 μL His-tag removed apo-$ACP_L$ (1 mM), 10 μL recombinant *Bacillus subtilis* Sfp (100 μM), 10 μL ATP (0.2 M), 25 μL terminal-$NH_2$ analogue of pantetheine (20 mM), 5 μL CoAA (1.2 mM) and 20 μL recombinant PPAT–DPCK (0.2 mM). The reaction was incubated at 37$^o$C for 2-3 hours. According to the HPLC analysis of the reaction mixture, 10-15% yield of $ACP_L$-$NH_2$ was achieved.

Two protocols were applied to purify $ACP_L$-NH2 from the reaction mixture. **Protocol A:** The reaction mixture was loaded onto a Source 15PRC reverse phase column (GE Healthcare). The protein were eluted with a linear gradient from 30% to 70% of $CH_3CN$ (0.1% $CF_3CO_2H$)/$H_2O$ (0.1% $CF_3CO_2H$). The $ACP_L$-$NH_2$ fractions was pooled, flash-frozen in liquid $N_2$ and lyophilized overnight. The lyophilized ACPL-$NH_2$ was redissolved and renatured in the storage buffer D for 1 hour before the assays. Compared with protocol B, protocol A has a faster purification and less loss of the protein, but only part of lyophilized $ACP_L$-$NH_2$ can be renatured. **Protocol B:** The reaction mixture was loaded onto HisTrap column to trap most of the His-tagged CoAA, PPAT–DPCK and Sfp, and the flow-through were pooled, concentrated, diluted 1:5 with 50 mM PBS buffer F (pH 8.0, 10% glycerol) and loaded onto 6 mL Resource Q ion-exchange column (GE Healthcare). $ACP_L$-$NH_2$ was eluted using a NaCl gradient 50-500 mM over 20 column volumes. The purified $ACP_L$-$NH_2$ was desalted by PD-10 column (GE Healthcare) equilibrated with storage buffer D, concentrated and checked for purity by SDS-PAGE and HPLC. Both protocol A and B can completely remove Sfp, which can interfere with the $GNAT_L$ loading assays.

**HPLC analysis of $GNAT_L$ decarboxylation of malonyl-CoA.** HPLC analysis of the decarboxylation was performed using XBridge C18 column (4.6 x 250 mm, 5 μm, Waters) on the Gold HPLC system equipped with an autosampler and controlled by 32 Karat software (Beckman Coulter). The samples were eluted with a linear gradient from 3% to 60% of MeOH/$H_2O$ (10 mM $CH_3CO_2NH_4$). For the steady-state kinetic studies, the

decarboxylations were examined in 40 μL 50 mM PBS, pH 7.0 at 100 μM, 200 μM, 500 μM, 1000 μM and 2000 μM malonyl or methylmalonyl-CoA. The $GNAT_L$ concentrations were optimized as 0.11 μM WT, 9.65 μM H389A, 14.6 μM H389N or 12.6 μM T355V for malonyl-CoA decarboxylation, and 0.35 μM WT for methylmalonyl-CoA decarboxylation. The reactions were incubated at room temperature for 5 min before quenched by 40 μL 1M $CH_3CO_2H$. 4 μL propionyl-CoA (for malonyl-CoA decarboxylation) or 4 μL acetyl-CoA (for methylmalonyl-CoA decarboxylation) was added as the internal standard. The reaction mixtures were filtrated by Microcon YM-10 (Millipore), neutralized by 20 μL 1M NaOH, and stored at -80$^o$C before analyzed by HPLC. Control reactions without enzymes were run at the same time. The peak areas of the decarboxylation products were normalized based on the internal standards, subtracted by the peak areas of the control reactions, and converted to turnovers/second based on peak areas of the acetyl or propionyl-CoA standard.

**Radioassays of $GNAT_L$ decarboxylation of malonyl-$ACP_L$.** [1,3-$^{14}$C]malonyl-CoA (55 mCi/mmol) were diluted with malonyl-CoA to get 5 mM 20 mCi/mmol [1,3-$^{14}$C]malonyl-CoA. [1,3-$^{14}$C]malonyl-$ACP_L$ was generated by incubating 100 μM apo-$ACP_L$ with 1 M [1,3-$^{14}$C]malonyl-CoA (20 mCi/mmol), 4 μM Sfp, 10 mM $MgCl_2$ in 50 mM Tris-HCl buffer, pH 8.1 at 30$^o$C for 2 h (*30*). The reaction mixture was desalted by PD-10 desalting column equilibrated with storage buffer D and concentrated. The final concentration of [1,3-$^{14}$C]malonyl-$ACP_L$ was measured by Bradford assay. The steady-state kinetic studies were performed in 20 μL 50 mM PBS, pH 7.0, 1.1 μM WT $GNAT_L$, at 15 μM, 30 μM, 60 μM, 120 μM and 180 μM malonyl [1,3-$^{14}$C]malonyl-$ACP_L$. The reactions were incubated at room temperature for 5 min before quenched with 60 μL 30% trichloroacetic acid (TCA) and 60 μL 10 mg/ml bovine serum albumin (BSA). The proteins were precipitated by centrifugation, washed twice with 10% TCA, redissolved in 100 μL formic acid, added into 4 mL scintillation fluid and counted. The amounts of radioactivity for the reactions were compared with those of controls and the decreased radioactivity were converted to turnovers/second by using [1,3-$^{14}$C]malonyl-CoA as the standard.

**Radioassays of AR-$GNAT_L$-$ACP_L$ acyl transfer.** Acyl transfer of AR-$GNAT_L$-$ACP_L$

was measured by using acetyl-CoA and (holo) AR-$GNAT_L$-$ACP_L$ (or the corresponding mutants). The kinetic studies were performed in 20 μL 50 mM MOPS, pH 7.0, 1 mM DTT, 25 μM (holo) AR-$GNAT_L$-$ACP_L$ (or the corresponding mutants), at 1 μM, 2.5 μM, 5 μM, 12.5 μM, 25 μM, 50 μM and 125 μM [1-$^{14}$C]acetyl-CoA. The (holo) $ACP_L$ was used in the control reactions to measure the autocatalytic transthioesterification. The reactions were incubated at room temperature for 1 min before quenched with TCA. The bound radioactivity of AR-$GNAT_L$-$ACP_L$ and $ACP_L$ was measured as above. The bound radioactivity of AR-$GNAT_L$-$ACP_L$ was subtracted by that of $ACP_L$, and converted to turnovers/second by using [1-$^{14}$C]acetyl-CoA as the standard. The acyl transfer rate was analyzed by Cleland's method (*31*).

**Autoradiographic analyses.** The loading of $GNAT_L$-$ACP_L$ or (DEBS) $AT_L$-$ACP_L$ was examined by SDS-PAGE followed by autoradiographic analyses. [2-$^{14}$C]malonyl-CoA (55 mCi/mmol) and [1-$^{14}$C]acetyl-CoA (55 mCi/mmol) were used for AR-$GNAT_L$-$ACP_L$ and $GNAT_L$-$ACP_L$ loading, and [1-$^{14}$C]propionyl-CoA (55 mCi/mmol) was used for (DEBS) $AT_L$-$ACP_L$ loading. The acyl transfer were performed in 50 mM MOPS, pH 7.0, 1 mM TCEP, 90 μM acyl-CoA, and 30 μM (holo or apo) AR-$GNAT_L$-$ACP_L$, $GNAT_L$-$ACP_L$, or (DEBS) $AT_L$-$ACP_L$. The aliquots were periodically moved from the reaction mixture and the proteins were precipitated by TCA, washed by ice-cold acetone and dissolved in Tris/glycine/SDS loading buffer. After running the SDS-PAGE, the gels were dried and subjected to autoradiography**.**

**Analysis of $ACP_L$ and AR-$GNAT_L$-$ACP_L$ digestion samples by ESI-FTICR-MS**. The observed and calculated masses for all the samples were listed in table 2-S4. $ACP_L$ samples were prepared by loading the reaction mixtures on Source 15PRC reverse phase column. The proteins were eluted with a linear gradient from 30% to 70% $CH_3CN$ (0.05% HCOOH and 0.05% $CF_3COOH$)/$H_2O$ (0.05% HCOOH and 0.05% $CF_3COOH$). AR-$GNAT_L$-$ACP_L$ digestion samples were prepared by first removing the CoA thioesters from AR-$GNAT_L$-$ACP_L$ reaction mixtures using Zeba desalt spin column (Pierce) equilibrated with the digestion buffer (50 mM $NH_4HCO_3$, pH 7.6, 0.1 M NaCl and 10% glycerol). 10 μL 2.5 mg/mL AR-$GNAT_L$-$ACP_L$ sample was digested with 5 μL 0.1

mg/mL trypsin (Princeton Separations) at 37$^{o}$C for 4 hours. The digestion was quenched by 0.1% formic acid and desalted by Source 15PRC column. (Performed by Liangcai Gu)

Mass spectrometric analysis was performed with an actively shielded 7 Tesla quadrupole-Fourier transform ion cyclotron resonance mass spectrometer (APEX-Q, Bruker Daltonics, Billerica, MA). Target analytes were diluted in an electrospray solution (1:1 $CH_3CN$:$H_2O$ with 0.1% HCOOH) and directly infused into an electrospray ionization (ESI) source (Apollo II, Bruker Daltonics) at a flow rate of 70 μL/h and a voltage of - 3.8 kV. A counterflow of hot (240 °C) nitrogen gas was applied to assist desolvation of ESI droplets. For accurate mass determination, up to 10 picomoles of apomyoglobin (Sigma, St. Louis, MO) was spiked into the ESI solution as internal calibrant. Multiply protonated ions generated from ESI were externally accumulated in a hexapole for 1 s and transferred via high voltage ion optics to the ICR cell for analysis. All data were acquired with XMASS software (version 6.1, Bruker Daltonics) in broadband mode from m/z = 200 to 2000 with 512k data points and summed over 20-30 scans. Mass spectra were analyzed with the MIDAS analysis software (*32*). When needed, external frequency-to-m/z calibration was performed with a two-term calibration equation (*33*) using two calibration standards (m/z = 622.02895 and 922.00979, from the calibration mix G2421A, Agilent Technologies, Palo Alto, CA). For infrared multiphoton dissociation (IRMPD), precursor ions were mass-selectively accumulated in a hexapole with a 3-5 m/z quadrupole isolation window, transferred to the ICR cell, and irradiated for 100-300 ms by 10.6 μm photons at 10 W laser power (25 W $CO_2$ laser, Synrad, Mukilteo, WA) for 30-50 scans. (Performed by Bo Wang)

**$GNAT_L$ preparation and crystallization.** In addition to pET24b::$GNAT_L$, pMCSG7::$GNAT_L$ was also constructed to facilitate His-tag cleavage (*25*). Selenomethionine-labeled $GNAT_L$ was produced in BL21(DE3) using SeMet minimal media according to the protocol reported (*34*). The Ni-affinity purification was performed in a similar way as mentioned above. The His-tag removal was performed in buffer F (20 mM Tris pH 7.9, 500 mM NaCl, 20 mM imidazole, 10% glycerol, and 1 mM DTT) by incubating $GNAT_L$ with 2% (w/w) His-tagged TEV protease at 4°C for 24 hr. The cleavage mixture was loaded onto HisTrap column, and flow-through fractions were

pooled and concentrated. After Ni-affinity purification (or His-tag cleavage), the concentrated fractions were loaded onto a HiLoad 16/60 Superdex 75 (GE Healthcare) column equilibrated with buffer G (20 mM Tris pH 7.9, 500 mM NaCl, and 10% glycerol). Fractions were then combined, concentrated to 13 mg/mL and either flash frozen in liquid $N_2$ or stored at 4°C. Crystals were grown in 24-72 hr at 20°C by micro-seeding in hanging drops using the vapor diffusion method. Equal volumes were mixed of protein solution and mother liquor containing 100 mM Bis-Tris pH 6.8, 100 mM NaCl, 1.0-1.5 M ammonium sulfate, and 20% glycerol. Apo crystals were harvested in loops and directly frozen by plunging into liquid $N_2$. In order to obtain the complex structure, apo crystals were grown at 4°C. After crystal growth, malonyl-CoA was added to the drop to a final concentration of ~10 mM and incubated for 1 hour before crystal mounting and freezing. (Performed by Todd Geders)

**Diffraction Data Collection and Structure Determination.** Diffraction data were collected at 100 K on GM/CA-CAT beamline 23ID-B at the Advanced Photon Source (APS) in Argonne National Laboratory (Argonne, IL). Data were processed using the HKL2000 suite (*35*). Initial phasing was performed using a three-wavelength MAD dataset from a single SeMet-labeled protein crystal of His-tagged $GNAT_L$. SOLVE was used to find the six selenium sites (‹m› = 0.52, score = 37.15) and for MAD phasing (overall FOM = 0.53) (*36*). RESOLVE was used for density modification (overall FOM = 0.66) and partial automated model building (*37, 38*). Two CurA molecules were present in the asymmetric unit. Modeling was completed manually using COOT (*39*). Refinement was performed using REFMAC5 of the CCP4 suite with TLS (*40-42*). The refined model was used as a probe structure for molecular replacement using PHASER with data from malonyl-CoA soaked crystals (*43, 44*). OMIT density was calculated using the program SFCHECK (*45*). Images and Figures were prepared using PyMOL (*46*). Tunnels were identified by the program CAVER (*47*). (Performed by Todd Geders)

**Modeling of Acetyl-CoA and ACP PPant Arm into Active Site.** Initial atomic coordinates and topology files for acetyl-CoA and 4'-phosphopantetheine were generated using the PRODRG2 server (*48*). Acetyl-CoA was positioned into the 1.95Å structure using the position of the low-occupancy conformation observed in the malonyl-CoA

soaked crystal as a guide. A fully extended 4'-phosphopantetheine was placed within the ACP binding tunnel. For both ligands, any overlapping waters were removed before energy minimization of an isolated CurA $GNAT_L$ molecule after the addition of hydrogens using the program CNS (*49*) in 500 steps of conjugate gradient minimization using no experimental energy terms and full harmonic restraints. (Performed by Todd Geders)

## 2.6 SUPPLEMENTARY FIGUES AND TABLES

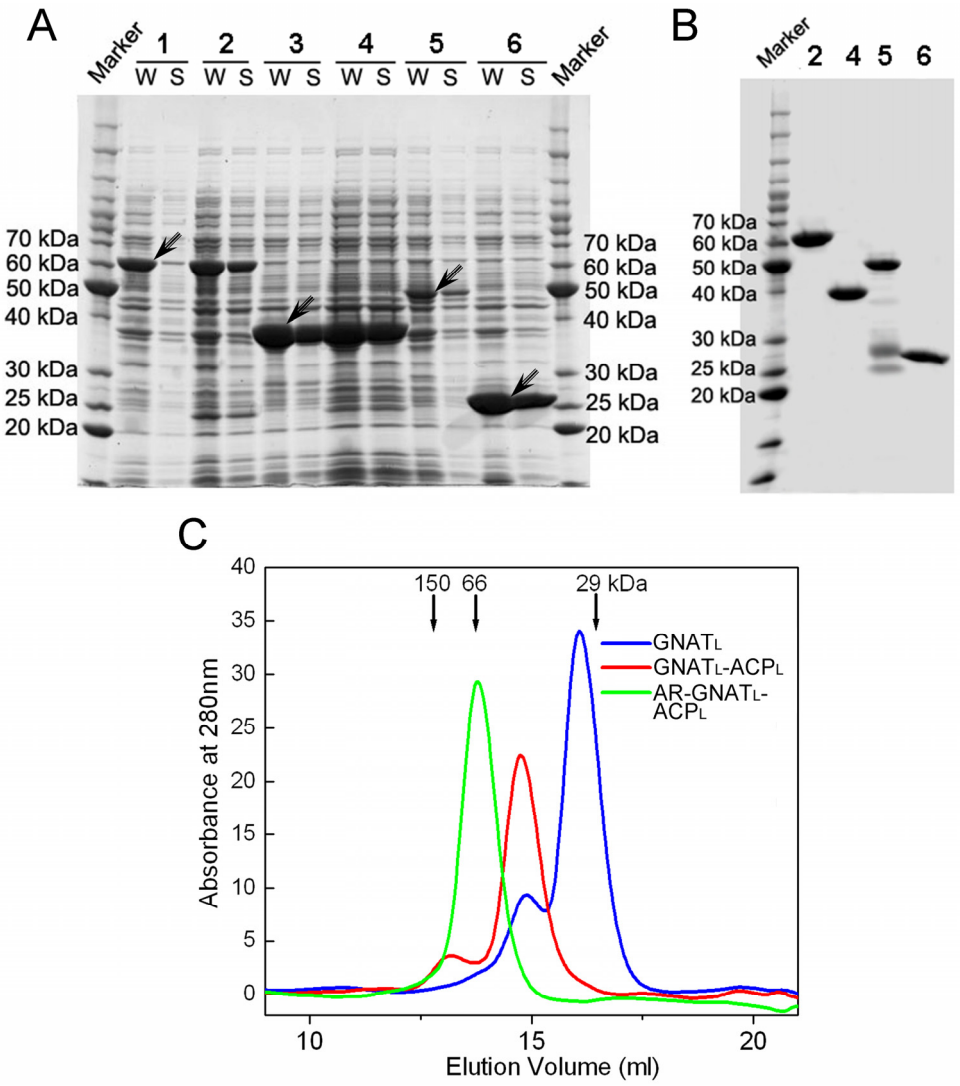



**Figure 2-S1. Protein expression, purification, and size-exclusion analysis.** (**A**) Expression and solubility of $GNAT_L$ constructs. Protein expression was induced at 20°C for 12 hours. W., whole cell lysate; S., supernatant. (**B**) Constructs after purification. 1. (apo) AR-$GNAT_L$-$ACP_L$, 2. (holo) AR-$GNAT_L$-$ACP_L$, 3. (apo) $GNAT_L$-$ACP_L$, 4. (holo) $GNAT_L$-$ACP_L$, 5. AR-$GNAT_L$, 6. $GNAT_L$. (**C**) Analytical size-exclusion chromatography of $GNAT_L$ constructs. 200 µL ~5mg/mL proteins were loaded on Superdex 200 10/300 GL column and eluted by 50 mM PBS, pH 7.5, 0.2 M NaCl and 10% glycerol. Partial dimerization of $GNAT_L$ and $GNAT_L$-$ACP_L$ was observed, however AR-$GNAT_L$-$ACP_L$ eluted exclusively as a monomer.

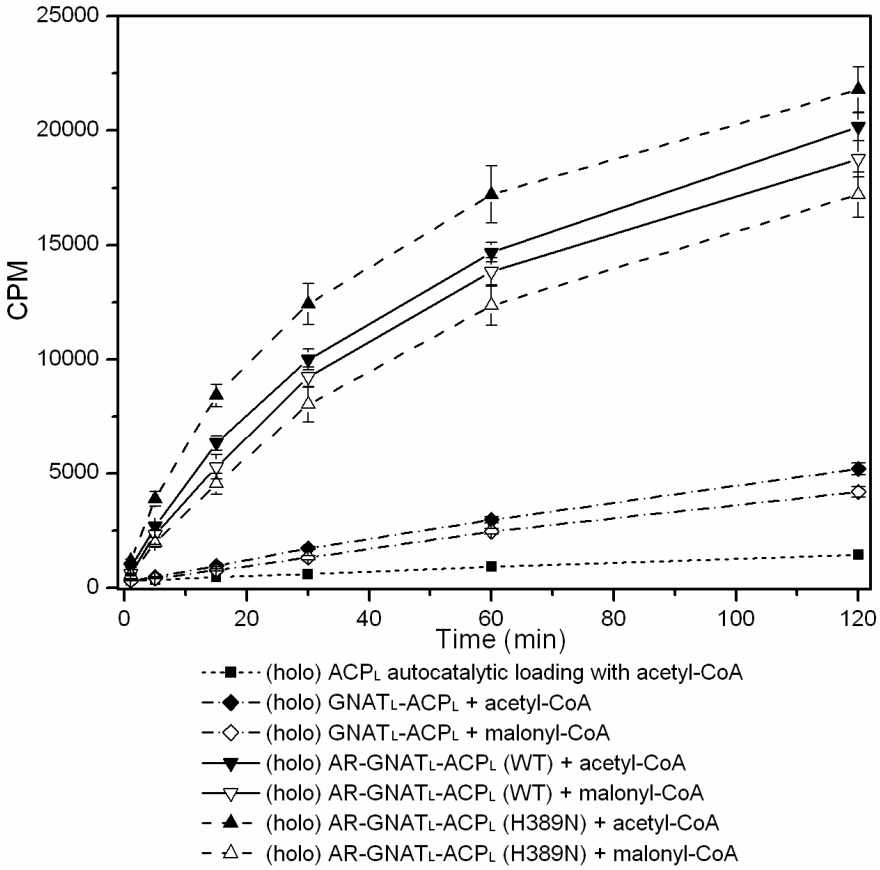



**Figure 2-S2. Acyl transfer *in cis* for (holo) AR-GNAT$_L$-ACP$_L$ (WT and H389N) and GNAT$_L$-ACP$_L$.** 20 μM holo enzymes were incubated with 100 μM [2-$^{14}$C]malonyl-CoA or [1-$^{14}$C]acetyl-CoA in 50 mM MOPS, pH 7.0, at room temperature. The autocatalytic acyl-CoA loading rate was measured using excised (holo) ACP$_L$ (20 μM). The H389N substitution resulted in ~3-fold greater $k_{cat}$ (see table S1) for acetyl transfer in comparison to wild type, perhaps by creating a stabilizing polar environment for the CoA thiolate intermediate generated immediately after acetyl transfer.

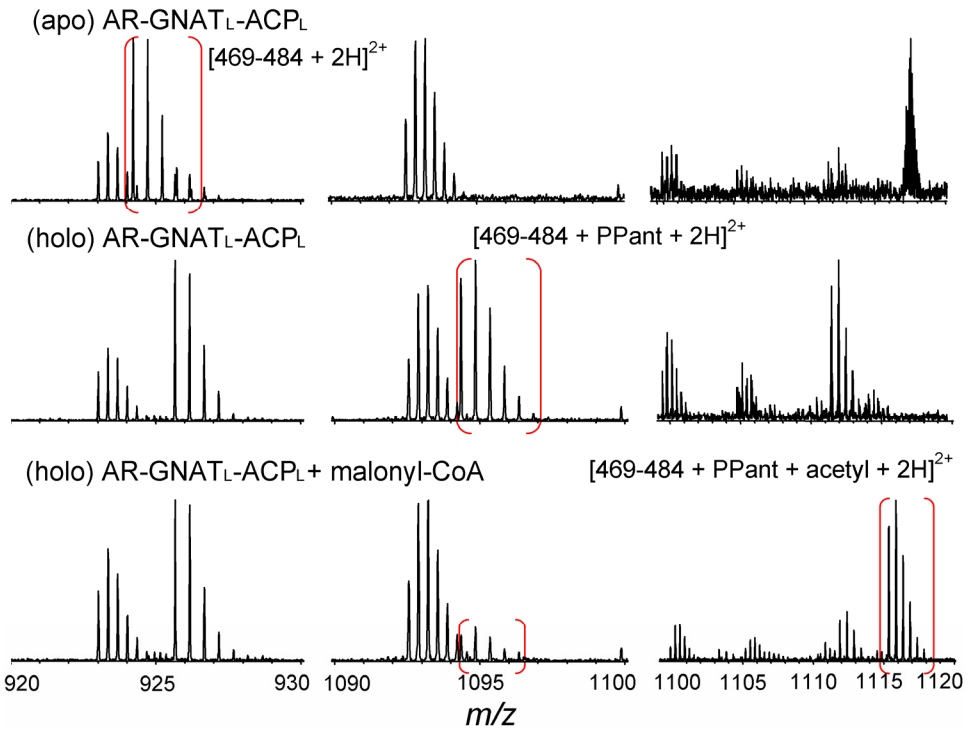



**Figure 2-S3.** Partial FTMS spectra showing the mass changes of serine$^{477}$-containing peptide from the trypsin digested (apo or holo) AR-GNAT$_L$-ACP$_L$ and malonyl-CoA loaded (holo) AR-GNAT$_L$-ACP$_L$.

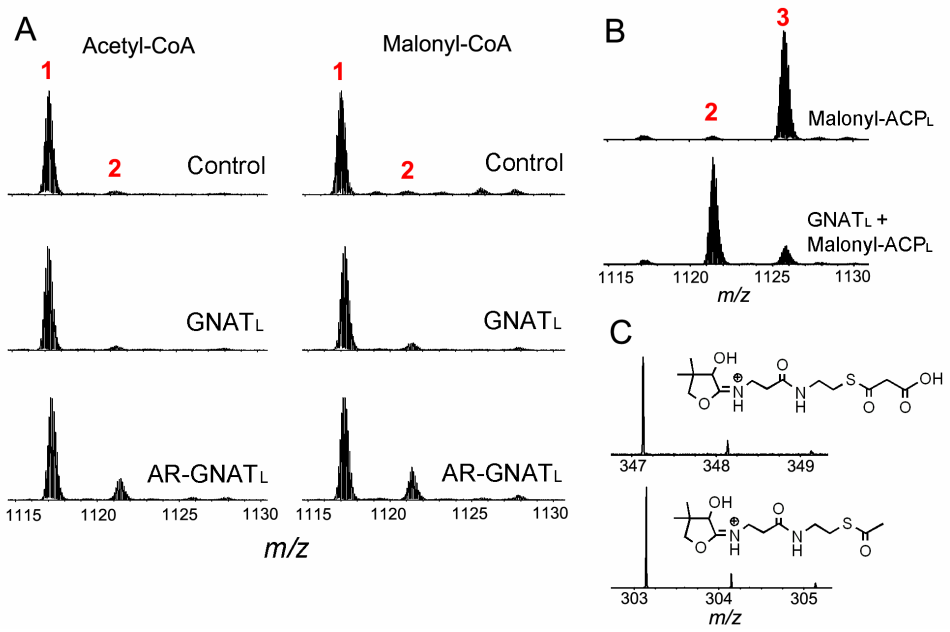



**Figure 2-S4.** (**A**) Comparison of the *in trans* loading activities of AR-$GNAT_L$ and $GNAT_L$ for acetyl and malonyl-CoAs. 15 μM (holo) $ACP_L$ was incubated with or without 2 μM AR-$GNAT_L$ or $GNAT_L$ in 50 μM CoAs and 50 mM MOPS, pH 7.0 at room temperature for 30 min. (**B**) $GNAT_L$ catalyzed decarboxylation of malonyl-$ACP_L$. 50 μM malonyl-$ACP_L$ was incubated with or without 1 μM $GNAT_L$ in 50 mM PBS, pH 7.0 at room temperature for 15 min. (**C**) IRMPD spectra showing the PPant ejection products for malonyl-$ACP_L$ (above) and acetyl-$ACP_L$ (below). **1**: $ACP_L$-SH, **2**: acetyl-$ACP_L$, **3**: malonyl-$ACP_L$.

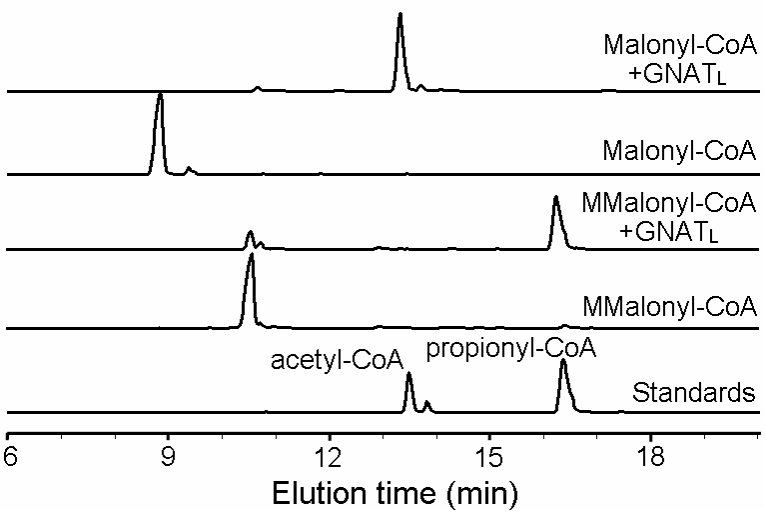



**Figure 2-S5. HPLC analysis of $GNAT_L$ catalyzed decarboxylation.** 200 μM malonyl-CoA/methylmalonyl-CoA and 1 μM $GNAT_L$ were incubated in 50 mM PBS, pH 7.0, at room temperature for 30 min. MMalonyl-CoA, methylmalonyl-CoA.

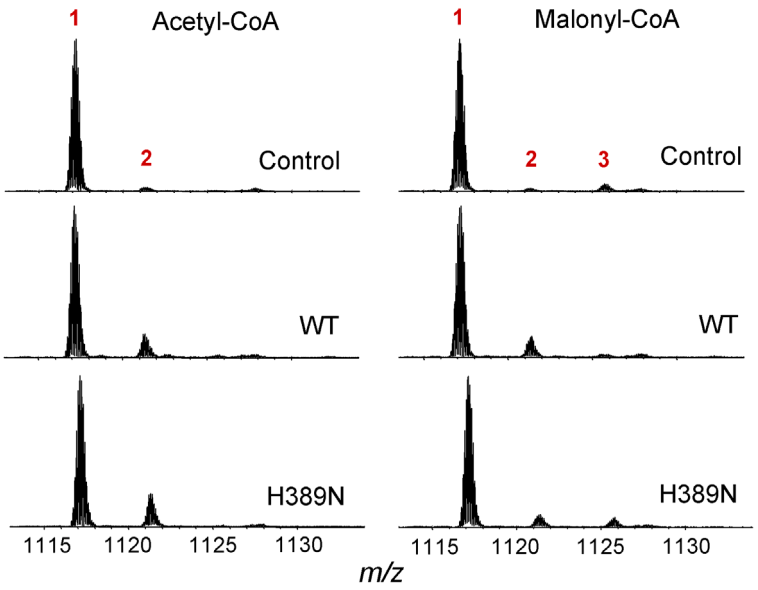
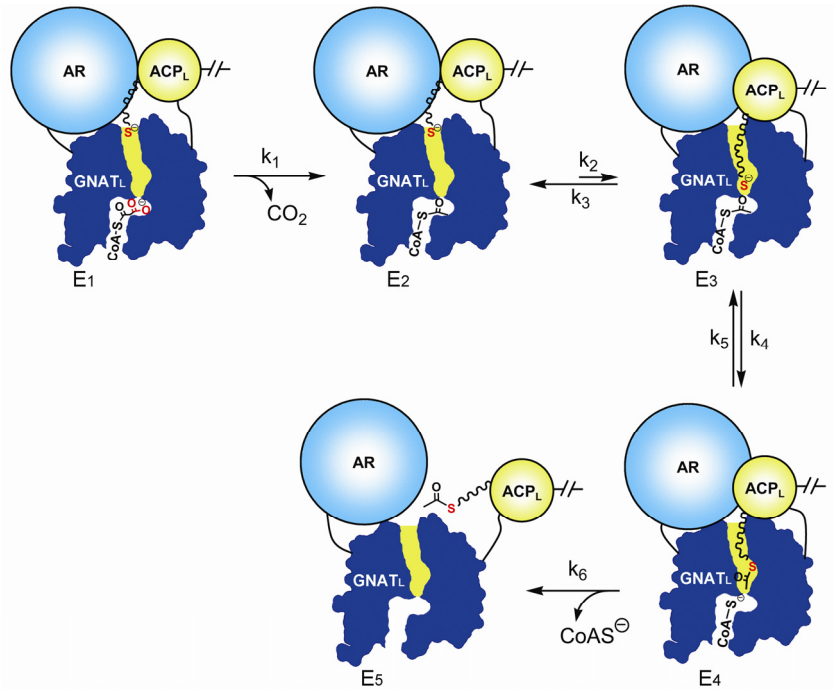



**Figure 2-S6. Comparison of the *in trans* loading activities of the wild type and H389N mutant of AR-GNAT$_L$-ACP$_L$.** 20 μM (holo) ACP$_L$ was incubated with 2 μM AR-GNAT$_L$-ACP$_L$ (WT or H389N) in 100 μM CoAs and 50 mM MOPS, pH 7.0 at room temperature for 30 min. **1**: ACP$_L$-SH, **2**: acetyl-ACP$_L$, **3**: malonyl-ACP$_L$. Malonyl-ACP$_L$ products (**3**) were observed for the malonyl-CoA loading by H389N mutant, which is likely due to the autocatalytic loading without decarboxylation.

**Figure 2-S7. Partition analysis of the acetyl transfer $k_{cat}$ of AR-GNAT$_L$-ACP$_L$ (holo) by using Cleland's method.**




A chain initiation mechanism for AR-GNAT$_L$-ACP$_L$ is proposed above. This mechanism can be reduced to mechanism (1) using net rate constants:

$$E_1 \xrightarrow{k_1'} E_2 \xrightarrow{k_2'} E_3 \xrightarrow{k_4'} E_4 \xrightarrow{k_6'} E_5 \quad (1)$$

The net rate constants are

$$k_1' = k_1,\ k_2' = \frac{k_2 k_4 k_6}{k_3 k_5 + k_3 k_6 + k_4 k_6},\ k_4' = \frac{k_4 k_6}{k_5 + k_6},\ k_6' = k_6 \quad (2)$$

The sum of [$E_4$] and [$E_5$] was measured for the acetyl transfer rate, and thus the acetyl transfer $k_{cat}$ can be calculated as:

$$k_{cat} = \frac{1}{\frac{1}{k_1} + \frac{1}{k_2'} + \frac{1}{k_4'}} \rightarrow \frac{1}{\frac{1}{k_2'} + \frac{1}{k_4'}} \text{ (assume } k_1 >> k_2' \text{ according to assays)}$$
$$= \frac{k_2 k_4}{k_2\left(1 + \frac{k_5}{k_6}\right) + k_3\left(1 + \frac{k_5}{k_6}\right) + k_4} \quad (3)$$

The net rate of acetyl transfer is negatively affected by (a) the slow rate of acetyl transfer ($k_4$), (b) the partitioning of $E_4$ between product release and reversal, and (c) the partitioning of $E_3$ between $E_2$ and $E_4$. Low occupancy of the GNAT$_L$ ACP tunnel ($k_2 < k_3$) can significantly decrease the acetyl transfer rate and the apparent $K_M$ of CoA substrates for acetyl transfer.

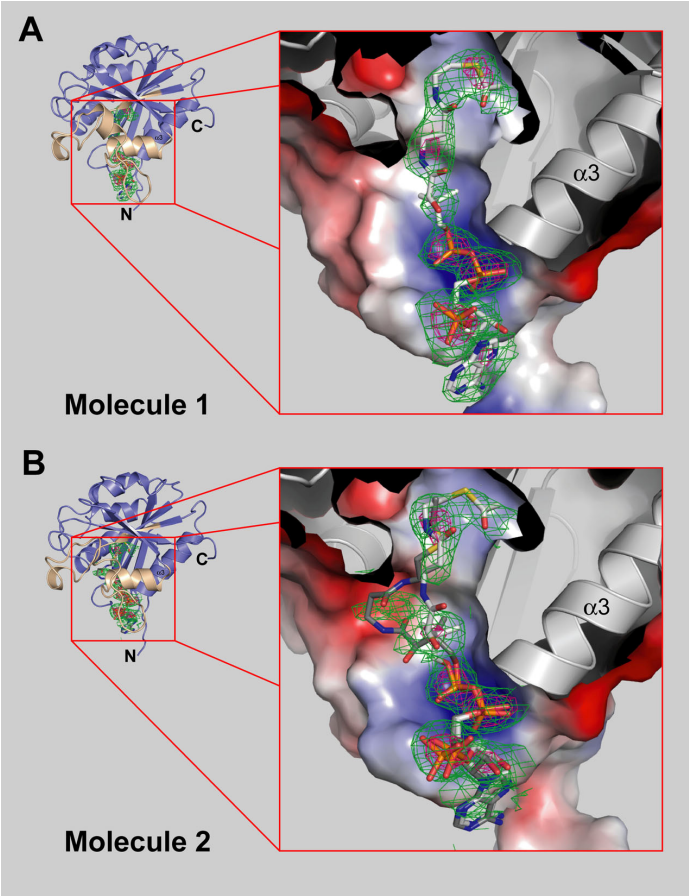



**Figure 2-S8. OMIT map density surrounding the CoA binding tunnel.** (**A**) First CurA $GNAT_L$ molecule in the asymmetric unit (ASU) with acetyl-CoA model drawn in white carbons. (**B**) Second CurA $GNAT_L$ molecule in the ASU with two alternate conformations of acetyl-CoA drawn in white and grey carbons. N- and C-termini are labeled for reference. Tan coloring indicates portions of the molecule omitted for clarity in inset. Protein surface is colored by electrostatic potential. OMIT density calculated from the final refined model is contoured at 1 σ (green) and 3 σ (magenta). Clear density observed for the 3'-ribose phosphate, which interacts with $Arg^{387}$ (not drawn), and the diphosphate linker, which is coordinated by backbone amides of residues 327-332 at the N-terminus of helix 3.

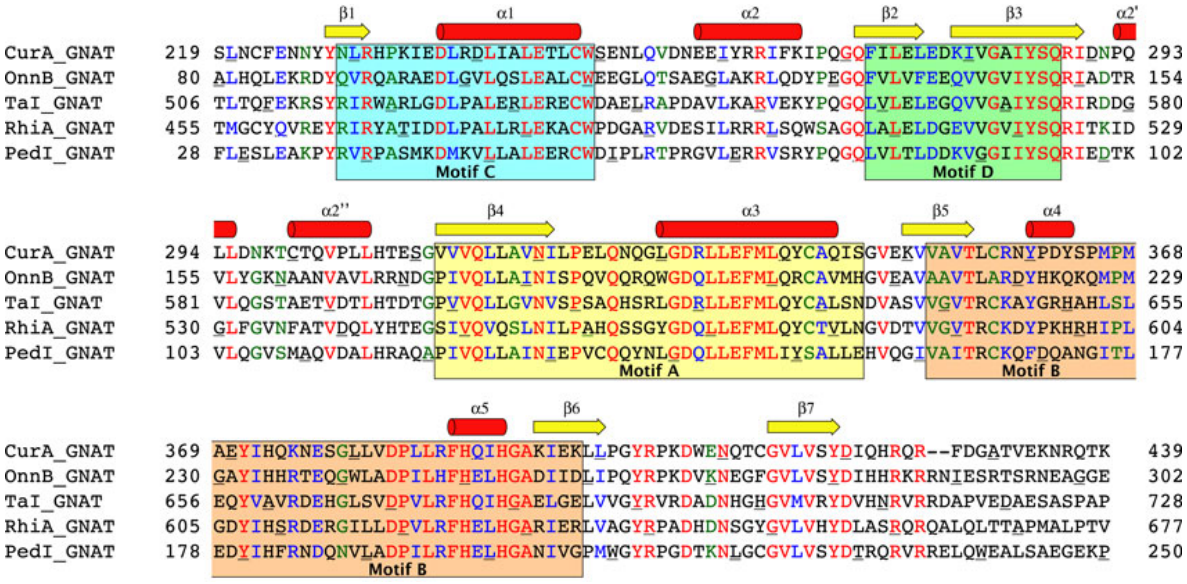



**Figure 2-S9.** Multiple sequence alignment of $GNAT_L$ domains from curacin A (CurA_GNAT), onnamide A (OnnB_GNAT, 47.5% identity), myxovirescin A (TaI_GNAT, 44.8% identity), rhizoxin (RhiA_GNAT, 43.4% identity), and pederin (PedI_GNAT, 40.7% identity) PKS clusters. Secondary structure elements of CurA $GNAT_L$ are shown above the alignment, colored boxes indicate regions designated motifs C, D, A, and B by Neuwald & Landsman, invariant residues are red, sites of conservative substitution are blue, and similar residues are green. Every tenth residue is underlined.

**Table 2-S1. Kinetic parameters for decarboxylation and acyl transfer.**

Decarboxylation

| $GNAT_L$ | Substrate | $k_{cat}$ [$s^{-1}$] | $K_M$ [μM] | $k_{cat}/K_M$ [x$10^3$ $M^{-1}s^{-1}$] |
|---|---|---|---|---|
| WT | malonyl-CoA | 1.8 ± 0.2 | 340 ± 27 | 5.25 ± 0.15 |
| H389A | malonyl-CoA | 0.018 ± 0.001 | 238 ± 15 | 0.077 ± 0.003 |
| H389N | malonyl-CoA | 0.025 ± 0.004 | 1460 ± 140 | 0.017 ± 0.002 |
| T355V | malonyl-CoA | 0.0071 ± 0.002 | 225 ± 21 | 0.0031 ± 0.005 |
| WT | methylmalonyl-CoA | 0.29 ± 0.07 | 168 ± 46 | 1.73 ± 0.07 |
| WT | malonyl-ACP | 0.037 ± 0.011 | 41 ± 13 | 0.92 ± 0.21 |

Acyl transfer

| (holo) AR-$GNAT_L$-$ACP_L$ | Substrate | $k_{cat}$ [$s^{-1}$] | $K_M$ [μM] | $k_{cat}/K_M$ [x$10^3$ $M^{-1}s^{-1}$] |
|---|---|---|---|---|
| WT | acetyl-CoA | 0.0023 ± 0.0002 | 4.3 ± 0.2 | 0.53 ± 0.02 |
| H389A | acetyl-CoA | 0.0019 ± 0.0003 | 4.2 ± 0.3 | 0.46 ± 0.03 |
| H389N | acetyl-CoA | 0.0067 ± 0.0005 | 14.0 ± 0.3 | 0.48 ± 0.03 |
| T355V | acetyl-CoA | 0.0031 ± 0.0008 | 5.7 ± 0.5 | 0.55 ± 0.07 |
| WT | malonyl-CoA | 0.0021 ± 0.0004 | 3.7 ± 0.6 | 0.58 ± 0.02 |

**Table 2-S2. Data collection and Refinement Statistics.**

**Table S2A – Data Collection**

| Parameter | Apo | Peak | Inflection | Remote | Malonyl-CoA Soak |
|---|---|---|---|---|---|
| Space Group | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ | $P3_121$ |
| Dimensions (Å) *a, c* | 91.5, 138.9 | 91.7, 137.5 | 91.7, 137.5 | 91.6, 137.5 | 91.9, 139.5 |
| X-ray source | 23ID-B | 23ID-B | 23ID-B | 23ID-B | 23ID-B |
| Wavelength λ (Å) | 0.97939 | 0.97934 | 0.97945 | 0.96112 | 0.97939 |
| $d_{min}$ (Å) | 1.95 | 2.30 | 2.30 | 2.30 | 2.75 |
| Unique observations | 49,929 | 30,361 | 30,365 | 30,327 | 18,244 |
| $R_{merge}$(%)[a,b] | 6.9 (55.5) | 7.2 (65.5) | 7.1 (63.7) | 7.2 (66.6) | 9.4 (53.4) |
| ‹I/σ›[a] | 28.9 (5.0) | 13.5 (1.5) | 14.0 (1.5) | 13.6 (1.5) | 24.5 (4.9) |
| Completeness (%)[a] | 99.6 (99.2) | 100 (100) | 100 (99.9) | 100 (100) | 99.9 (99.9) |
| Avg. redundancy[a] | 11.2 (11.1) | 2.9 (2.9) | 2.9 (2.8) | 2.9 (2.9) | 9.7 (9.8) |

[a]Values in parenthesis are for outer shell

[b]$R_{merge}=\sum|I_i-\langle I\rangle|/\sum I_i$, where $I_i$ is the intensity of the *i*th observation and ‹I› is the mean intensity

**Table S2B – Refinement Statistics**

Refinement statistics

| | Apo | Malonyl-CoA soak |
|---|---|---|
| Data range | 50-1.95 | 50-2.75 |
| $R/R_{free}$[a] | 0.172/0.211 | 0.207/0.251 |
| RMSD bond length (Å) | 0.013 | 0.008 |
| RMSD bond angle (Å) | 1.382 | 1.351 |
| Avg. Protein B-factor (Å$^2$) | 39.5 | 43.1 |
| Avg. Solvent B-factor (Å$^2$) | 47.4 | 28.8 |
| Wilson B (Å$^2$) | 31.6 | 60.3 |
| Ramachandran plot[c] | | |
| Favored | 97.7 | 96.4 |
| Allowed | 3.3 | 3.4 |
| Disallowed | 0.0 | 0.2 |
| Protein atoms | 3432 | 3356 |
| Water molecules | 460 | 4 |
| Other atoms | 33 | 153 |
| PDB code | 2REE | 2REF |

[a]$R = \sum|F_O-|F_C||/\sum|F_O|$ where $F_O$ is the observed structure factor and $F_C$ is the calculated structure factor used in the refinement

[b]$R_{free} = \sum|F_O-|F_C||/\sum|F_O|$ where $F_O$ is the observed structure factor and $F_C$ is the calculated structure factor from 5% of reflections not used in the refinement

[c]From output of MOLProbity (Davis *et al.* 2004 Nucleic Acids Research 32:W615-W619)

**Table 2-S3. Primers for the expression constructs and site mutagenesis.**

| Construction Primers | |
|---|---|
| Primer name | Sequences (5' to 3') |
| AR-GNAT$_L$-ACP$_L$-for | CATATGCTAGAGTTAATTAATCGT (*Nde*I) |
| AR-GNAT$_L$-ACP$_L$-rev | CTCGAGTTACTCTGGCTTTGGTAGTAC (*Xho*I) |
| GNAT$_L$-ACP$_L$-for | CATATGAGTTTAAATTGTTTTGAAAATAAT (*Nde*I) |
| GNAT$_L$-ACP$_L$-rev | CTCGAGTTACTCTGGCTTTGGTAGTAC (*Xho*I) |
| AR-GNAT$_L$-for | CATATGCTAGAGTTAATTAATCGT (*Nde*I) |
| AR-GNAT$_L$-rev | CTCGAGCTTAGTCTGCCGATTTTTTTC (*Xho*I) |
| GNAT$_L$-for | CATATGAGTTTAAATTGTTTTGAAAATAAT (*Nde*I) |
| GNAT$_L$-rev | CTCGAGCTTAGTCTGCCGATTTTTTTC (*Xho*I) |
| ACP$_L$-for | CATATGAAACTGACAGAAGATATAGAT (*Nde*I) |
| ACP$_L$-rev | CTCGAGTTACTCTGGCTTTGGTAGTAC (*Xho*I) |
| AR-for | CATATGCTAGAGTTAATTAATCGT (*Nde*I) |
| AR-rev | CTCGAGTTTTTGGACTTTTATCAAAC (*Xho*I) |
| (DEBS)AT$_L$-ACP$_L$-for | CATATGGCGGACCTGTCAAAGC (*Nde*I) |
| (DEBS)AT$_L$-ACP$_L$-rev | CTCGAGGCGGGTTTCCCGTTGTGC (*Xho*I) |
| GNAT$_L$7-for (LIC) | TACTTCCAATCCAATGCCAGTTTAAATTGTTTTGAAAATAATTATTACAATTTGCGCCATCC |
| GNAT$_L$7-rev (LIC) | TTATCCACTTCCAATGCTACTTAGTCTGCCGATTTTTTTCAACTGTCGC |
| **Mutagenic Primers** | |
| Primer name | Sequences (5' to 3') |
| H389A-for | GGATCCGCTGTTGCGATTTGCCCAGATTCATGGGGCTAAAATTG |
| H389A-rev | CAATTTTAGCCCCATGAATCTGGGCAAATCGCAACAGCGGATCC |
| H389N-for | GGATCCGCTGTTGCGATTTAACCAGATTCATGGGGCTAAAATTG |
| H389N-rev | CAATTTTAGCCCCATGAATCTGGTTAAATCGCAACAGCGGATCC |
| T355V-for | GGTGTCGAAAAAGTTGTGGCAGTAGTTCTTTGTCGCAATTATCCAG |
| T355V-rev | CTGGATAATTGCGACAAAGAACTACTGCCACAACTTTTTCGACACC |
| S477A-for | GATGGAGATGGGCATCGAAGCTTTGGAGTTGCTGGAAC |
| S477A-rev | GTTCCAGCAACTCCAAAGCTTCGATGCCCATCTCCATC |

The restriction sites are underlined; the mismatch bases are in red.

**Table 2-S4. MS analysis for decarboxylation and acetyl transfer.**

| *$ACP_L$ samples* | | | | |
|---|---|---|---|---|
| $ACP_L$* | ESI-FT-ICR-MS | | IRMPD (PPant ejection product) | |
| | Obs. avg mass | Calc. avg mass | Obs. $[M + H]^+$ | Calc. $[M + H]^+$ |
| apo-$ACP_L$ | 10820.7 | 10821.1 | | |
| holo-$ACP_L$ | 11160.9 | 11160.7 | 216.128 | 261.127 |
| acetyl-S-$ACP_L$ | 11202.5 | 11202.7 | 303.138 | 303.137 |
| malonyl-S-$ACP_L$ | 11247.4 | 11247.2 | 347.128 | 347.127 |
| $NH_2$-$ACP_L$ | 11144.5 | 11144.7 | 244.166 | 244.166 |
| Acetyl-NH -$ACP_L$ | 11186.4 | 11186.7 | ND | ND |
| *AR-$GNAT_L$-$ACP_L$ trypsin digestion samples* | | | | |
| $[469\text{-}484 + 2H]^{2+}$ | ESI-FT-ICR-MS | | IRMPD (PPant ejection product) | |
| | Obs. mass | Calc. mass | Obs. $[M + H]^+$ | Calc. $[M + H]^+$ |
| apo- | 1845.97 | 1845.95 | | |
| holo- | 2186.05 | 2186.04 | 261.127 | 261.127 |
| acetyl- | 2228.06 | 2228.05 | 303.138 | 303.137 |

Masses are in Daltons. Calc., calculated; Obs., observed; ND, not detected

*The masses ACP derivatives are calculated without the first methionine for the $His_6$-tag removed protein.

†[469-484] = ALMEMGIES$^{477}$LELLELR.

§PPant ejection product for acetyl-NH-ACP was not detected due to the weak signal of acetyl-NH-$ACP_L$.

## 2.7. REFERENCES


1. C. Bisang *et al.*, *Nature* **401**, 502 (1999).
2. Y. Q. Xue, L. S. Zhao, H. W. Liu, D. H. Sherman, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12111 (1998).
3. J. Lau, D. E. Cane, C. Khosla, *Biochemistry* **39**, 10514 (2000).
4. Y. J. Hsieh, P. E. Kolattukudy, *J. Bacteriol.* **176**, 714 (1994).
5. P. Verdier-Pinard *et al.*, *Mol. Pharmacol.* **53**, 62 (1998).
6. Z. X. Chang *et al.*, *J. Nat. Prod.* **67**, 1356 (2004).
7. J. Piel, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14002 (2002).
8. J. Piel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16222 (2004).
9. V. Simunovic *et al.*, *Chembiochem* **7**, 1206 (2006).
10. L. P. Partida-Martinez, C. Hertweck, *Chembiochem* **8**, 41 (2007).
11. F. Dyda, D. C. Klein, A. B. Hickman, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 81 (2000).
12. M. W. Vetting *et al.*, *Arch. Biochem. Biophys.* **433**, 212 (2005).
13. C. Khosla, Y. Y. Tang, A. Y. Chen, N. A. Schnarr, D. E. Cane, *Annu. Rev.B iochem.* **76**, 11.1 (2007).
14. S. Smith, A. Witkowski, A. K. Joshi, *Prog. Lipid Res.* **42**, 289 (2003).
15. T. Maier, S. Jenni, N. Ban, *Science* **311**, 1258 (2006).
16. Y. Y. Tang, C. Y. Kim, Mathews, II, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11124 (2006).
17. L. E. N. Quadri *et al.*, *Biochemistry* **37**, 1585 (1998).
18. P. C. Dorrestein *et al.*, *Biochemistry* **45**, 12756 (2006).
19. R. H. Lambalot *et al.*, *Chem. Biol.* **3**, 923 (1996).
20. A. B. Hickman, D. C. Klein, F. Dyda, *Mol. Cell* **3**, 23 (1999).
21. R. M. Van Wagoner, J. Clardy, *Structure* **14**, 1425 (2006).
22. J. V. Rossi, M. A. Roberts, H. D. Yoo, W. H. Gerwick, *J. Appl. Phycol.* **9**, 195 (1997).
23. W. Minas, P. Brunker, P. T. Kallio, J. E. Bailey, *Biotechnol. Progr.* **14**, 561 (1998).
24. J. L. Meier, A. C. Mercer, H. Rivera, M. D. Burkart, *J. Am. Chem. Soc.* **128**, 12174 (2006).
25. L. Stols *et al.*, *Prot. Expres. Purif.* **25**, 8 (2002).
26. B. S. Kim *et al.*, *Biochemistry* **41**, 10827 (2002).
27. M. Daugherty, V. Vonstein, R. Overbeek, A. Osterman, *J. Bacteriol.* **183**, 292 (2001).
28. A. S. Worthington, M. D. Burkart, *Org. Biomol. Chem.* **4**, 44 (2006).

29. K. F. Geoghegan *et al.*, *Anal. Biochem.* **267**, 169 (1999).
30. L. C. Gu *et al.*, *J. Am. Chem. Soc.* **128**, 9014 (2006).
31. W. W. Cleland, *Biochemistry* **14**, 3220 (1975).
32. M. W. Senko, J. D. Canterbury, S. H. Guan, A. G. Marshall, *Rapid Commun. Mass Spectrom.* **10**, 1839 (1996).
33. E. B. Ledford, D. L. Rempel, M. L. Gross, *Anal. Chem.* **56**, 2744 (1984).
34. S. A. Guerrero, H. J. Hecht, B. Hofmann, H. Biebl, M. Singh, *Appl. Microb. Biotechnol.* **56**, 718 (2001).
35. Z. Otwinowski, W. Minor, in *Macromol. Crystallogr. A* (1997), vol. 276, pp. 307-326.
36. T. C. Terwilliger, J. Berendzen, *Acta Crystallogr. D* **55**, 849 (1999).
37. T. C. Terwilliger, *Acta Crystallogr. D* **56**, 965 (2000).
38. T. C. Terwilliger, *Acta Crystallogr. D* **59**, 38 (2003).
39. P. Emsley, K. Cowtan, *Acta Crystallogr. D* **60**, 2126 (2004).
40. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. D* **53**, 240 (1997).
41. S. Bailey, *Acta Crystallogr. D* **50**, 760 (1994).
42. J. Painter, E. A. Merritt, *Acta Crystallogr. D* **62**, 439 (2006).
43. L. C. Storoni, A. J. McCoy, R. J. Read, *Acta Crystallogr. D* **60**, 432 (2004).
44. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr. D* **61**, 458 (2005).
45. A. A. Vaguine, J. Richelle, S. J. Wodak, *Acta Crystallogr. D* **55**, 191 (1999).
46. W. L. DeLano, *The PyMOL Molecular Graphics System*, http://www.pymol.org (2002).
47. M. Petrek *et al.*, *BMC Bioinform.* **7**, (2006).
48. A. W. Schuttelkopf, D. M. F. van Aalten, *Acta Crystallogr. D* **60**, 1355 (2004).
49. A. T. Brunger *et al.*, *Acta Crystallogr. D* **54**, 905 (1998).

**Notes:**

Liangcai Gu and David Sherman designed and performed the experiments;

Todd Geders and Janet Smith solved the $GNAT_L$ structure and designed the site-directed mutagenesis;

Bo Wang and Kristina Håkansson recorded and analyzed the FT-ICR mass spectra.

# CHAPTER 3
# POLYKETIDE HMG β-BRANCHING

## *3.1 SUMMARY*

This chapter contains two successive studies towards understanding the mechanisms of polyketide HMG β-branching in the curacin A pathway. The first study describes the functional identification of a pair of mechanistically diverse enzymes that catalyze the successive dehydration (CurE $ECH_1$) and decarboxylation (CurF $ECH_2$) of (S)-HMG-ACP to generate a 3-methylcrotonyl-ACP intermediate, the presumed precursor of the cyclopropyl ring in curacin A. The reactions catalyzed by $ECH_1$ and $ECH_2$ are found in a broad cross-section of microbial natural product gene clusters and participate in the introduction of carbon chain branch points and functional group diversity as key steps in the HMG-CoA synthase mediated addition of C-2 from acetate to the α-carbonyl group of polyketide chains. The second study provides structural insights into CurF $ECH_2$ decarboxylation in polyketide chain β-branching. The crystal structure of the CurF N-terminal $ECH_2$ domain establishes that the protein is a crotonase superfamily member. $Ala^{78}$ and $Gly^{118}$ form an oxyanion hole in the active site, which includes only three polar side chains as potential catalytic residues. Site-directed mutagenesis and a coupled $ECH_1/ECH_2$ dehydration/decarboxylation assay established critical functions for $His^{240}$ and $Lys^{86}$, whereas $Tyr^{82}$ was non-essential. A decarboxylation mechanism is proposed in which $His^{240}$ serves to stabilize the substrate carboxylate and $Lys^{86}$ donates a proton to γ-C of the acyl-ACP enolate intermediate to form the α,β unsaturated isopentenoyl-ACP product. CurF $ECH_2$ substrates were synthesized enzymatically, and used in a direct $ECH_2$ assay to show a 20-fold selectivity for ACP- over CoA-linked substrates. Specificity for ACP-linked substrates has not been reported for any other crotonase superfamily decarboxylase. $Tyr^{73}$ may select against

CoA-linked substrates by blocking a contact of $Arg^{38}$ with the CoA adenosine 5'-phosphate.

## 3.2 INTRODUCTION

Polyketides and non-ribosomal peptides are important secondary metabolites possessing an array of biological activities and considerable chemical diversity (*1-4*). Type I modular polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) systems sequentially elongate and modify a growing ketide or peptide chain as it is passed from module to module. The biosynthetic machinery generates tremendous chemical diversity in these compounds by use of specific reductive domain combinations within modules and/or gene cassettes that encode proteins capable of producing unique functionalities.

Curacin A (**1**, Figure 3-1A), a marine cyanobacterial metabolite from *Lyngbya majuscula*, is a mixed polyketide-non-ribosomal peptide natural product with potent anti-proliferative and cytotoxic activity against colon, renal, and breast-cancer derived cell lines (*5*). Because of its unusual structure, which includes a cyclopropane group, thiazoline moiety, *cis*-alkenyl group, and terminal double bond, we were motivated to characterize the unique enzymes responsible for the curacin A assembly.

Recently, we identified and characterized a gene cluster (*cur*) from *Lyngbya majuscula* L19 whose architecture and domain organization correlates directly with that expected for curacin A biosynthesis (*6*). Within the *cur* gene cluster that spans ~65 kbp is an ~8 kbp region (CurA-CurF) with striking similarity (80-95% amino acid identity) to a comparably-sized section of the jamaicamide **2** (*jam*) biosynthetic pathway (JamE-JamJ) (*7*). The genes from this region are translated into a set of enzymes containing a predicted α-ketoglutarate dependent halogenase (*8*), a set of three tandem ACPs, and a 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase (HCS)-like gene cassette that collectively are likely involved in the introduction of the cyclopropane group in curacin A and the vinyl chloride substituent in jamaicamide (*6, 7*). The HCS-like gene cassette encodes five biosynthetic enzymes: an acyl carrier protein (CurB), a putative ketosynthase (CurC), an HMG-CoA synthase (CurD), and two putative enoyl-CoA

hydratases (CurE $ECH_1$ and CurF $ECH_2$). This novel cassette has recently been identified in several additional gene clusters, including the *jam* pathway from *L. majuscula*, the mupirocin **3** (*Mup*) pathway from *Pseudomonas fluorescens*, and the "PksX" pathway

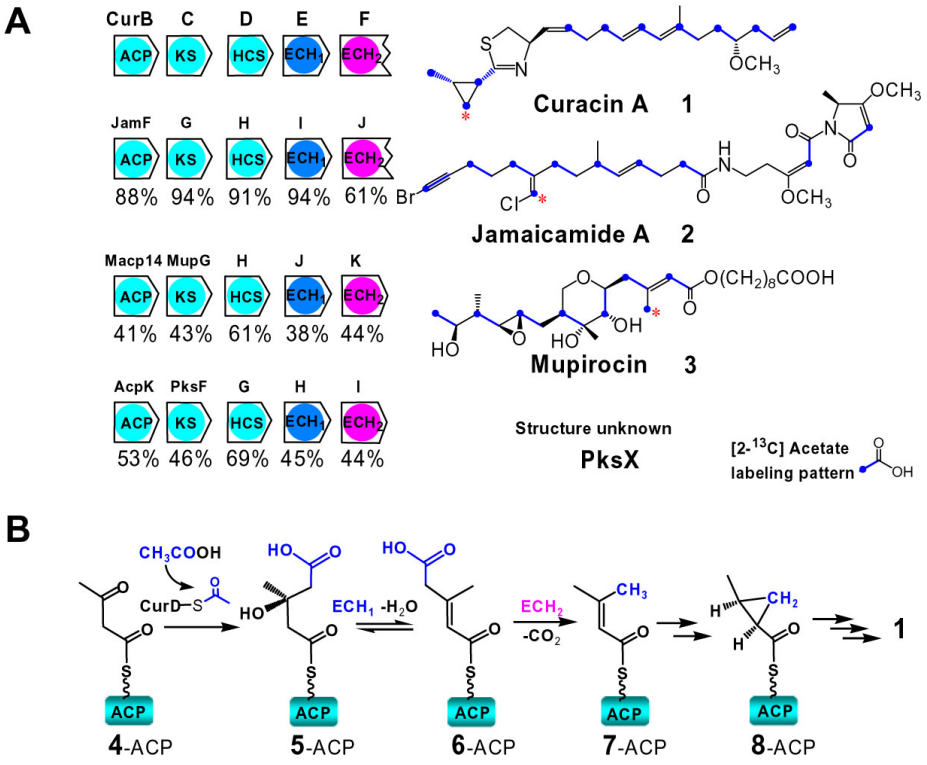



**Figure 3-1. HMG enzyme cassettes. A**) HMG gene cassettes are proposed to introduce a pendant carbon (indicated by red asterisk) from C-2 of acetate to the polyketide chains. Amino acid identities of enzymes are shown (%), and acetate labeling patterns are highlighted in blue. **B**) Proposed mechanism for the HMG β-branching in curacin A pathway.

from *Bacillus subtilis* (*6-9*), and is proposed to catalyze addition of C-2 from acetate onto the polyketide chain to generate a pendant functional group (Figure 3-1A).

In the first study (*10*), we provided evidence for the specific function of the CurE/CurF $ECH_1$-$ECH_2$ enzyme pair (Figure 3-1A) in curacin A biosynthesis, and demonstrated that they catalyze the successive dehydration and decarboxylation of (*S*)-HMG-ACP **5** to generate a 3-methylcrotonyl-ACP **7** intermediate for the subsequent formation of the cyclopropane ring in **8** (Figure 3-1B). Significantly, it is now evident from bioinformatic analysis that the $ECH_1$-$ECH_2$ enzyme pair exists widely in different

microbial species and is not necessarily limited to HCS-like gene cassettes (Figure 3-S1). Therefore, we believe that the $ECH_1$-$ECH_2$ of the curacin A system represents an excellent example of chemistry-directed enzyme evolution, giving rise to "mechanistic diversity"(*11*) within a single protein family.

Enoyl-CoA hydratases belong to the crotonase superfamily, consisting of a wide variety of mechanistically diverse enzymes that exhibit various activities, such as hydratase (*12*), dehalogenase (*13*), decarboxylase (*13, 14*), isomerase (*15-17*), hydrolase (*14*), and carbon-carbon bond formation (*14*) or cleavage (*18, 19*). The common mechanistic theme of this low-sequence-identity superfamily is the stabilization of an enolate anion intermediate of phosphopantetheine-linked substrates by two backbone amide groups forming an oxyanion hole (*20*). A unique feature of the $ECH_2$-like enzymes involved in natural product biosynthesis is that they are thought to depend exclusively on ACP-linked substrates *in vivo*. Structural clues that enable these enzymes to discriminate between CoA and ACP-linked substrates will increase our understanding of the evolutionary changes that have occurred to favor one substrate over another, as well as to control functional group diversity in the final natural product.

In the second study (*21*), we solved the 1.85Å crystal structure of the wild-type, N-terminal $ECH_2$ domain of CurF and the 1.65Å structure of the corresponding Y82F variant. Modeling of the substrate of CurF $ECH_2$ was performed and $Tyr^{82}$, $Lys^{86}$, and $His^{240}$ were identified as potential catalytic or substrate-binding residues. Site-directed mutagenesis in a coupled $ECH_1$/$ECH_2$ dehydration/decarboxylation assay with ACP-linked substrates demonstrated that CurF $ECH_2$ $His^{240}$ and $Lys^{86}$ are critical to catalysis.

## 3.3 RESULTS AND DISCUSSION

### 3.3.1 Functional Identification of $ECH_1$ and $ECH_2$ Enzyme Pair

In the *cur* gene cluster, $ECH_1$ is encoded by *curE*, and the $ECH_2$ gene is embedded in *curF* as the first domain of the CurF hybrid PKS/NRPS bimodular polypeptide (Figure 3-1A). Although the sequence identity between *cur* $ECH_1$ and $ECH_2$ is only 17%, they

are predicted to belong to the crotonase superfamily, and contain two consensus sequences (Figure 3-S1) essential for the "oxyanion hole" to stabilize enolate anions (*20*). In contrast, the sequence identities in the $ECH_1$ and $ECH_2$ groups compared from various biosynthetic systems (Figure 3-S1) are generally above 40%, which suggests that $ECH_1$ and $ECH_2$ have different functional roles. To identify their precise function, we reasoned that the HCS motif would follow a path involving initial assembly of HMG-ACP via the reaction of C-2 of acetyl-ACP with acetoacetyl-ACP **4** (catalyzed by HMG-CoA synthase (CurD)). The co-linearity of the Cur biosynthetic pathway suggested that the $ECH_1$-$ECH_2$ pair would operate in late stage construction of the presumed acyl-ACP precursor for cyclopropane ring formation.

To probe the function of $ECH_1$, *curE* was cloned, overexpressed and purified as a 30.5 kDa N-His-tagged fusion protein. Next, we sought to excise a functional form of $ECH_2$ from the amino terminus of the CurF multifunctional enzyme. This was accomplished by choosing an optimized cutting site in the linker region of $ECH_2$ (259 amino acids) to yield a 29.5 kDa C-His-tagged fusion protein. Additionally, *curB* was cloned, overexpressed and purified in the apo-CurB (acyl carrier protein) form as an 11 kDa N-His-tagged fusion protein. $ECH_1$, $ECH_2$ and apo-CurB were isolated in soluble form from *E. coli* and purified by Ni-NTA agarose resin. The apo-CurB protein (75 μM) was subsequently loaded using HMG-CoA via *in vitro* phosphopantetheinylation through *in situ* incubation with (*R,S*)-HMG-CoA (500 μM) and Sfp (4 μM) (*22*). The HMG-holo-CurB was dialyzed against ECH assay buffer to adjust pH and remove excess HMG-CoA.

The *in vitro* activity of $ECH_1$ and $ECH_2$ were investigated by incubating $ECH_1$, $ECH_2$ or both (2 μM each) with (*R,S*)-HMG-holo-CurB (50 μM) in a series of buffers at 37$^{o}$C. ESI-FTMS (Apex-Q instrument, Bruker Daltonics) was applied to detect mass change of the acyl group covalently linked to the holo-CurB. We found that in the presence of $ECH_2$ alone (Figure 3-2A, c), no new reaction products were observed. In contrast, in the presence of $ECH_1$, a peak corresponding to 18 Da loss in molecular mass occurred (Figure 3-2A, b), and when both $ECH_1$ and $ECH_2$ were employed, two peaks corresponding to 18 Da and 62 Da loss in molecular mass were observed (Figure 3-2A, d).

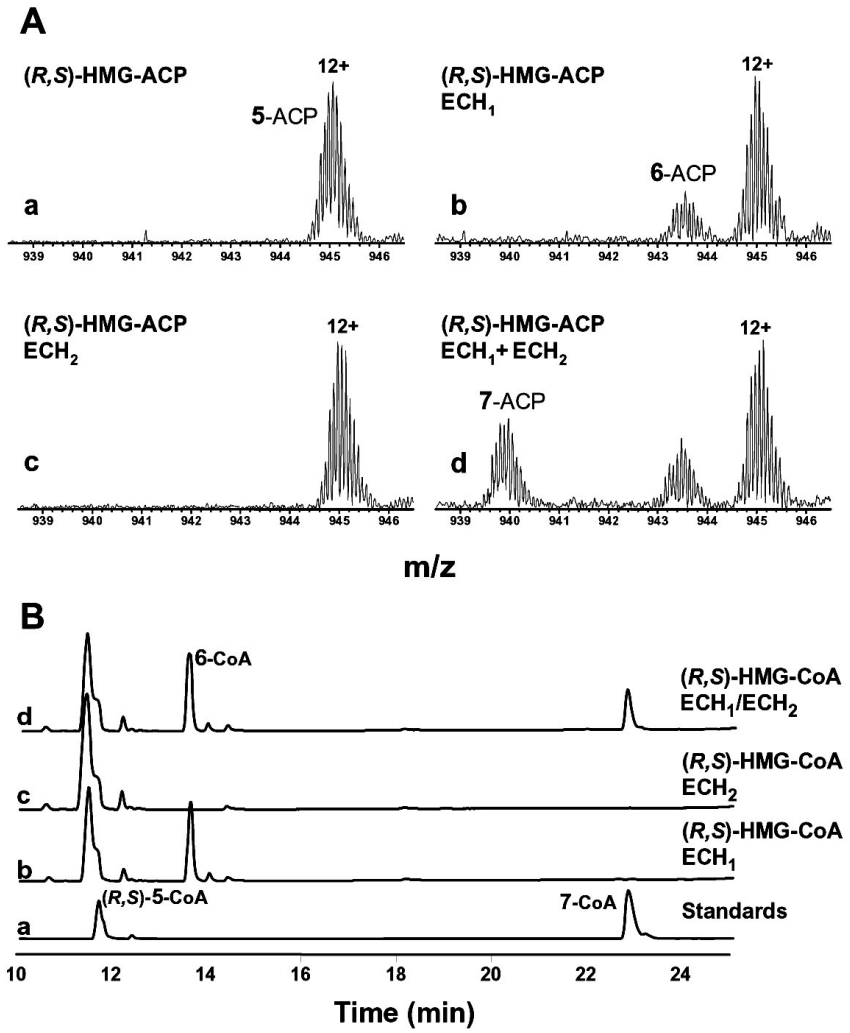



**Figure 3-2. $ECH_1$ and $ECH_2$ assays for the substrates in CoA and ACP forms. A**) FTICR spectra, 50 μM (*R*,*S*)-HMG-ACP, 2 μM $ECH_1$, $ECH_2$, or both, at 37°C for 3 h. Experimental most abundant mass: **5**-ACP, 11325.8; **6**-ACP, 11307.8; **7**-ACP, 11264.8. **B**) UV 275 nm traces of (a) standards: **5**-CoA and **7**-CoA (MC-CoA), 0.5 mM (*R*,*S*)-**5**-CoA treated with (b) 2 μM $ECH_1$, (c) 2 μM $ECH_2$, (d) 2 μM $ECH_1$ and $ECH_2$ at 37°C for 3 h. The CoA peak shoulders and the following minor peaks are possibly due to CoA aggregation.

These results suggest that $ECH_1$ catalyzes dehydration of HMG-ACP **5** to 3-methylglutaconyl-ACP **6** (Figure 3-1B), and $ECH_2$ catalyzes subsequent decarboxylation to **7.**

Next, we also demonstrated that $ECH_1$ and $ECH_2$ are able to accept and catalyze dehydration and decarboxylation of (*R,S*)-HMG-CoA (Figure 3-2B), which suggests that

the two enzymes recognize the phosphopantetheine arm from CoA as well as holo-ACP. Importantly, this finding facilitated structural identification of the dehydration and decarboxylation products from the reaction. Specifically, 0.5 mM (*R,S*) HMG-CoA was incubated with the $ECH_1$, $ECH_2$ or both (2 μM each) in 30 mM Bis-Tris buffer, pH 6.5 at 37 $^{o}$C for 3 hours, and the reaction mixtures were analyzed by HPLC equipped with a photodiode array detector and an ESI-LTQ mass spectrometer (ThermoFinnigan). UV spectral analysis of the HMG-CoA substrate, and the following dehydration and decarboxylation steps, revealed a significant absorption increase at 260 nm for both dehydration and decarboxylation products (Figure 3-S2). This increase is consistent with an α,β C=C bond formation in an acyl-CoA thioester (*23*). In addition, MS/MS analysis confirmed that the dehydration product was **6**-CoA, and we did not pursue the further clarification of the C=C configuration. To determine regiochemistry of the double bond in the decarboxylation product, we synthesized 3-methyl-3-butenoyl-CoA and used commercially available **7**-CoA (Sigma) as authentic standards. HPLC co-injection showed that the decarboxylation product is **7**-CoA (Figure 3-S3A).

Finally, we investigated the substrate preference of $ECH_1$ by comparing the conversion ratio of (*R,S*)-HMG-CoA and (*S*)-HMG-CoA which was generated by HMG-CoA reductase (*24*). Based on HPLC traces, the conversion ratio of (*S*)-HMG-CoA is 2-fold higher than that of (*R,S*)-HMG-CoA (Figure 3-S3B), which indicates that (*S*)-HMG is the natural $ECH_1$ substrate (*25*).

In summary, the CurE/CurF $ECH_1$-$ECH_2$ enzyme pair from the curacin A biosynthetic pathway was functionally identified as a mechanistically diverse enzyme pair. We demonstrated that CurE $ECH_1$ catalyzes dehydration of (*S*)-HMG-ACP **5** to form 3-methylglutaconyl-ACP **6**, and CurF $ECH_2$ catalyzes decarboxylation of **6** to generate 3-methylcrotonyl-ACP **7**, the presumed precursor for cyclopropyl-ACP **8** formation in curacin A. The detailed steps leading from 3-methylcrotonyl-ACP to the cyclopropane ring are the subject of ongoing studies in our laboratory. It is noteworthy that, to date, only two members of crotonase enzyme superfamily, methylmalonyl-CoA decarboxylase (MMCD) (*26*) and CarB (*27*) were reported to catalyze loss of carbon dioxide. Thus, identification of the reaction catalyzed by CurF $ECH_2$ provides a new

example of this novel biotin-independent decarboxylase in secondary metabolism.

### *3.3.2 Structural Insights into $ECH_2$ Decarboxylation*

**Structure Determination** Initial crystal screening with a polypeptide including residues 1-257 of CurF produced crystals that diffracted to only ~3.8Å, but an N-terminal truncation including residues 17-257 yielded crystals that diffracted to beyond 2Å. The structure was solved by MAD phasing using selenomethionyl CurF $ECH_2$ (Table 3-S1). The resulting model was refined against the 2.0Å dataset in crystal form I and used to solve the structures by molecular replacement in crystal form II (Table 3-S2).

**Structure of the N-terminal $ECH_2$ Domain of CurF** The CurF $ECH_2$ domain possesses the crotonase core fold (residues 17-209) consisting of a central mostly parallel beta sheet flanked by alpha helices (Figure 3-3A). Roughly perpendicular to the main beta sheet (β1, β2, β3, β5, β7) are short beta strands (β4, β6, β8) that are a hallmark of the crotonase superfamily. Crotonase superfamily members can be classified into one of three structural classes based on the conformation of the helical C-terminus (*14, 20*). The helical C-terminus (α9, α10, α11) of CurF $ECH_2$ has the "self-association" fold seen in at least eight other crotonase family members (*15-17, 19, 28*), including the two with biotin-independent decarboxylase activities, methylmalonyl CoA decarboxylase (MMCD) (*26*) and carboxymethylproline synthase (CarB) (*14*).

**Quaternary Structure of the $ECH_2$ Domain of CurF** The trimer structure fundamental to the crotonase superfamily occurs in both crystal forms of $ECH_2$ (Figure 3-3B). Subunit contacts are virtually identical in the two crystal forms. A total of 15% of the solvent-accessible surface area of the monomer is buried in the $ECH_2$ trimer (total buried surface area per monomer = 1050 $Å^2$). The extensive buried surface and the lack of water in the subunit interface together indicate that the trimeric association observed in the crystal structure reflects a true quaternary structure for the protein. However, in solution CurF $ECH_2$ also displayed concentration-dependent dissociation (Figure 3-S4), indicating a dynamic equilibrium between trimeric and lower oligomeric states.

**Active site** Despite less than 20% sequence identity with other crotonase family

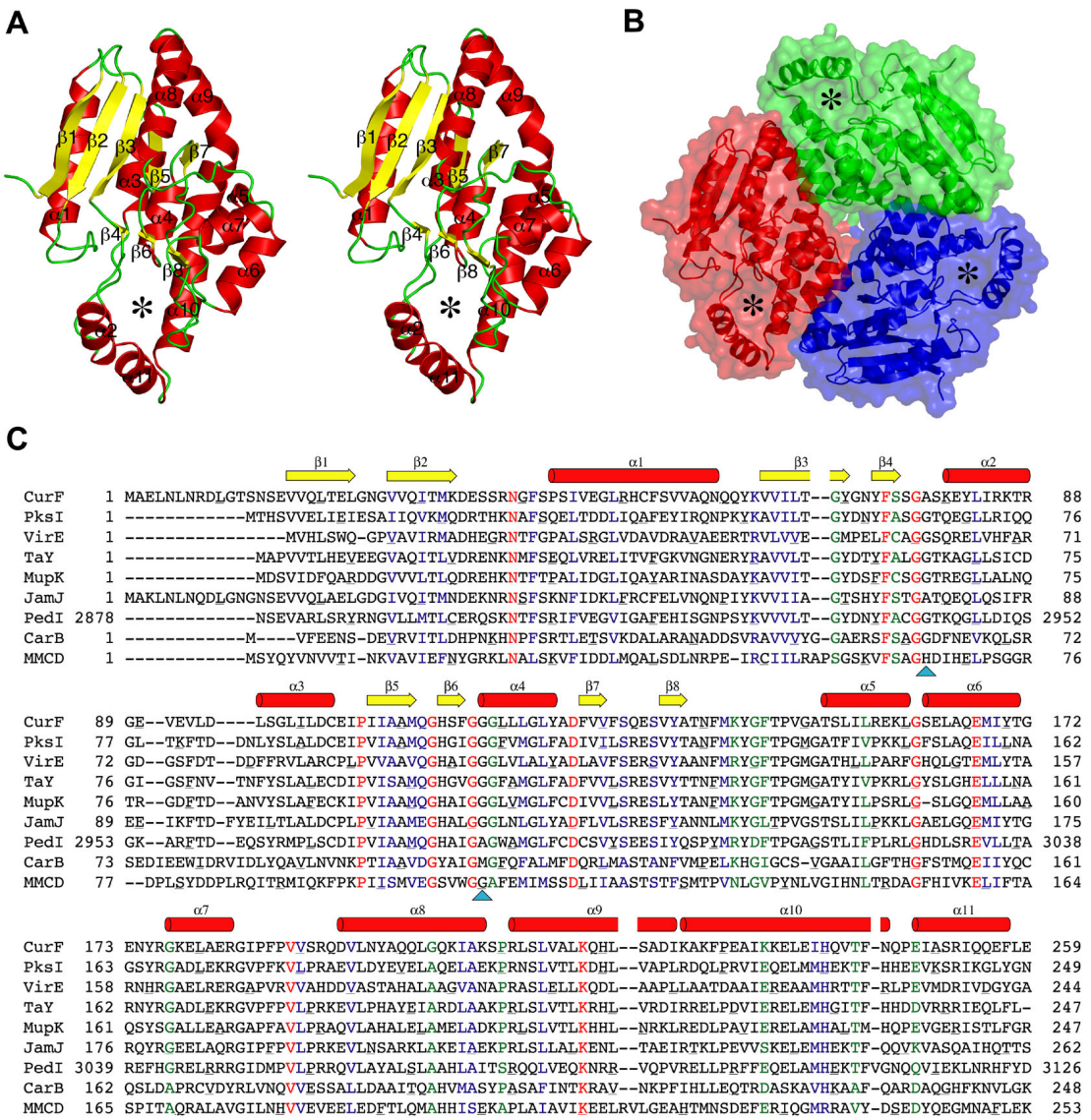



**Figure 3-3. CurF $ECH_2$ structure. A**), stereo diagram of the monomer of the CurF $ECH_2$ domain with the active site chamber indicated by an *asterisk.* **B**), trimeric structure of the CurF $ECH_2$ domain with the active site chamber indicated by an *asterisk.* **C**), multiple sequence alignment of $ECH_2$-like decarboxylases from gene clusters encoding the biosynthesis of curacin (CurF), bacillaene (PksI, 44% identity), virginiamycin M (VirE, 36% identity), myxovirescin A (TaY, 43% identity), mupirocin (MupK, 39% identity), jamaicamide (JamJ, 59% identity), pederin (PedI, 36% identity), and sequences of the two structurally characterized crotonase superfamily decarboxylases (CarB [17% identity] and MMCD [15% identity]). Secondary structure elements of CurF are shown above the alignment, residues of the oxyanion hole are indicated by *triangles* below the alignment, invariant residues are *red*, sites of conservative substitution are *blue*, and similar residues are *green*. Every tenth residue is *underlined*.aggregation.

members of known structure, strong structural similarity of the CurF $ECH_2$ domain to

crotonase superfamily members enabled us to align the structures and to identify critical elements of the active site, including the substrate binding tunnel, active site oxyanion hole, and active site chamber. Structural alignments with several other crotonase family members (Figure 3-S5) clearly indicate that the backbone amides of residues Ala[78] and Gly[118] form the oxyanion hole. These residues follow conserved Gly[77] and Gly[117], which have backbone conformations only accessible to glycine. In this manner, the peptide planes of residues 77-78 and 117-118 are oriented so their amides can stabilize the

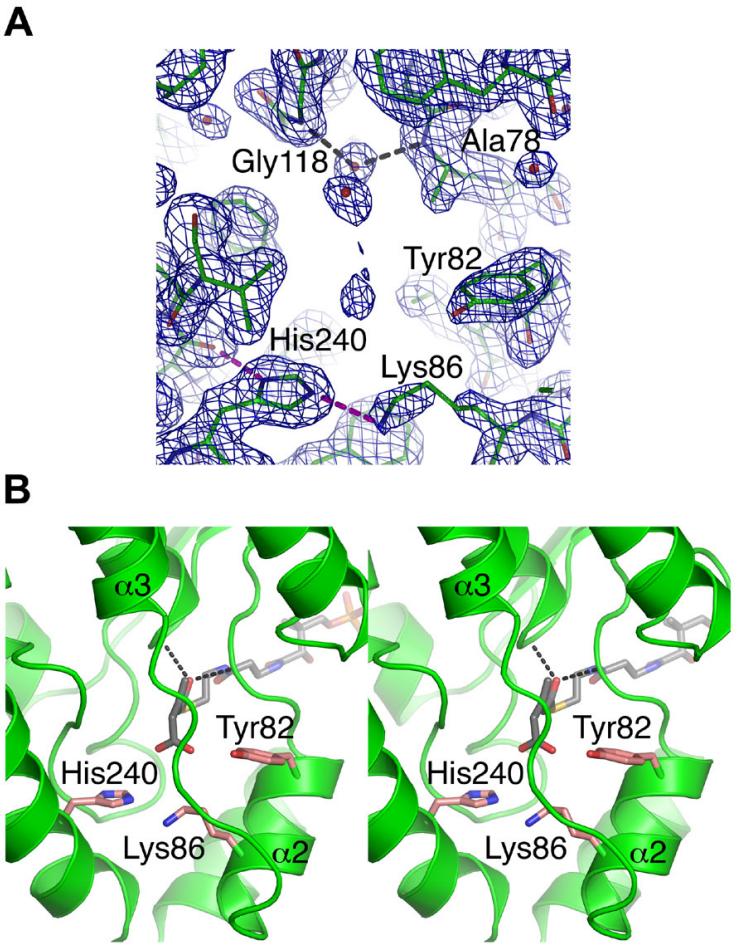



**Figure 3-4. Active site of CurF $ECH_2$. A**) refined $2F_o$ - $F_c$ electron density contoured at 1 σ from wild type CurF $ECH_2$ domain. The water molecule present in the oxyanion hole is shown with *dashed lines* to the backbone amides of Ala[78] and Gly[118]. Residues His[240], Lys[86], and Tyr[82] are shown in the foreground. **B**) stereo diagram of substrate (*gray*) modeling results highlighting polar residues within8Åofthe modeled substrate carboxylate group. The loop connecting helices α2 and α3 covers the active site.

proposed enolate anion intermediate by hydrogen bonding. In the crystal structures, a water molecule occupies the oxyanion hole (Figure 3-4A). Attempts to obtain crystal structures of complexes with product analogues were unsuccessful. The substrate of CurF $ECH_2$ was then modeled into the active site (see Methods). Only three polar side chains ($Tyr^{82}$, $Lys^{86}$ and $His^{240}$) are present within the primarily hydrophobic active-site chamber (Figure 3-4B).

**Decarboxylase Activity with ACP-linked substrate** Based on the active site and substrate modeling, residues $Tyr^{82}$, $Lys^{86}$ and $His^{240}$ were selected for mutagenesis. Proteins possessing Y82F, K86A, K86Q H240A, or H240Q were produced and assayed for activity against ACP-linked substrates in a coupled $ECH_1/ECH_2$ assay (*10*). Substitution of $Tyr^{82}$ by Phe led to two-fold reduction in product formation in comparison to wild-type (Figure 3-5), indicating that $Tyr^{82}$ plays little or no role in catalysis. On the other hand, product formation was reduced more than 20-fold in the K86A, K86Q, H240A, and H240Q variants compared to wild-type. FT-ICR mass spectrometry confirmed that the $ECH_1$ and $ECH_2$ products were produced from the ACP-linked substrate (Figure 3-5C, Figure 3-S6) (*10*). The 1.65Å crystal structure of the Y82F variant of CurF $ECH_2$ showed a similar position for the phenyl ring within an otherwise undisturbed structure. Although stable, the CurF $ECH_2$ K86A, K86Q, H240A, and H240Q variants have not yielded crystals.

**Decarboxylase Activity Preference for ACP-linked Substrates over CoA-linked Substrates** CurE $ECH_1$ activity towards (*R*,*S*)-HMG-CoA or (*R*,*S*)-HMG-ACP yielded very similar 3-methylglutaconyl-ACP and 3-methylglutaconyl-CoA yields (Figure 3-6). CurF $ECH_2$ activity towards 3-methylglutaconyl-ACP or 3-methylglutaconyl-CoA yielded significantly better activity with the ACP-linked substrate (~3% CoA conversion versus >70% conversion, respectively) (Figure 3-6).

The N-terminal $ECH_2$ domain of CurF possesses the crotonase superfamily fold. As is seen with the two other structurally characterized crotonase family members possessing biotin-independent decarboxylase activity (CarB (*14*) and MMCD (*26*)), the C-terminal helical domain folds back upon the monomer in a "self association" fold (*15*).

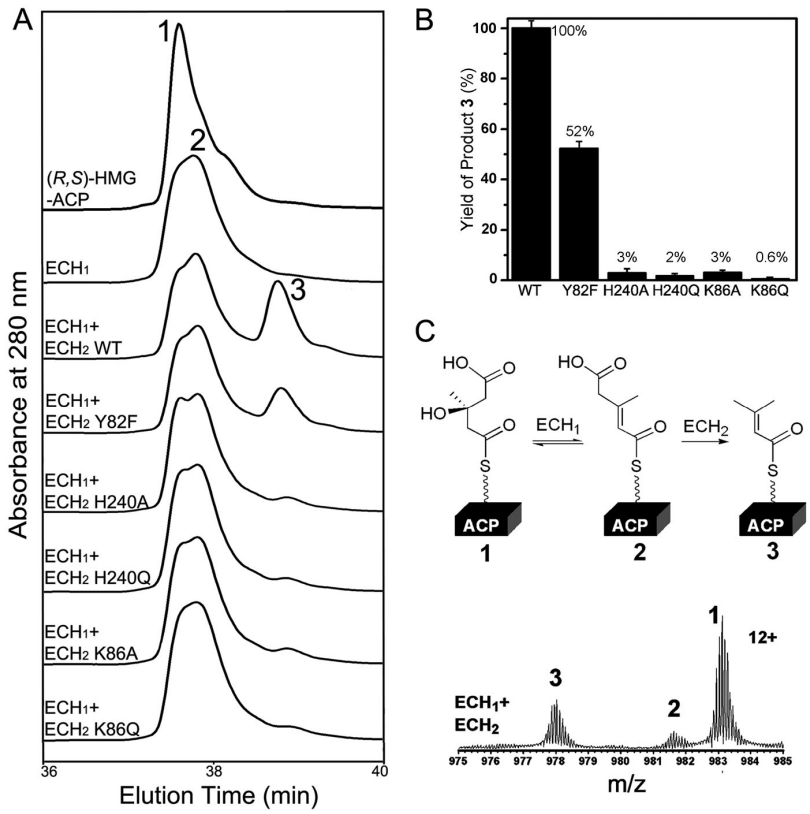



**Figure 3-5. $ECH_1/ECH_2$ coupled enzymatic assays for $ECH_2$ wild type and mutants. A**) HPLC chromatograms showing the (*R,S*)-HMG-CurA-ACP(II) and associated species before and after $ECH_1/ECH_2$ coupled reactions at 37°C for 1 h; the peaks correspond to (*R,S*)-HMG-CurA-ACP(II) (*peak 1*), 3-methylglutaconyl-CurA-ACP(II) (*peak 2*), and 3-methylcrotonyl-CurA-ACP(II) (*peak 3*). **B**) comparison of the peak areas of 3-methylcrotonyl-CurA-ACP(II) for $ECH_2$ wild type and mutants. The data were subjected to normalization and base-line subtraction before the peak area calculation, and the assays were duplicated. **C**) FTICR-MS identification of the ACP-linked substrate and products. Only the preferred substrate for $ECH_1$,(*S*)-HMG-ACP, is shown in the reaction scheme.

Thus the active site chamber is formed exclusively by residues within one monomer of the trimer.

The diversity of chemical reactions catalyzed by crotonase family members and their highly divergent sequences result in little if any conservation within the active site chamber. Nevertheless, the close structural similarity with other crotonase family members suggests strongly that $Ala^{78}$ and $Gly^{118}$ form the oxyanion hole of CurF $ECH_2$ (Figure 3-S5). The site-directed mutagenesis results show that $Lys^{86}$ and $His^{240}$ play crucial roles in catalysis (Figure 3-5). We propose a decarboxylation mechanism in which

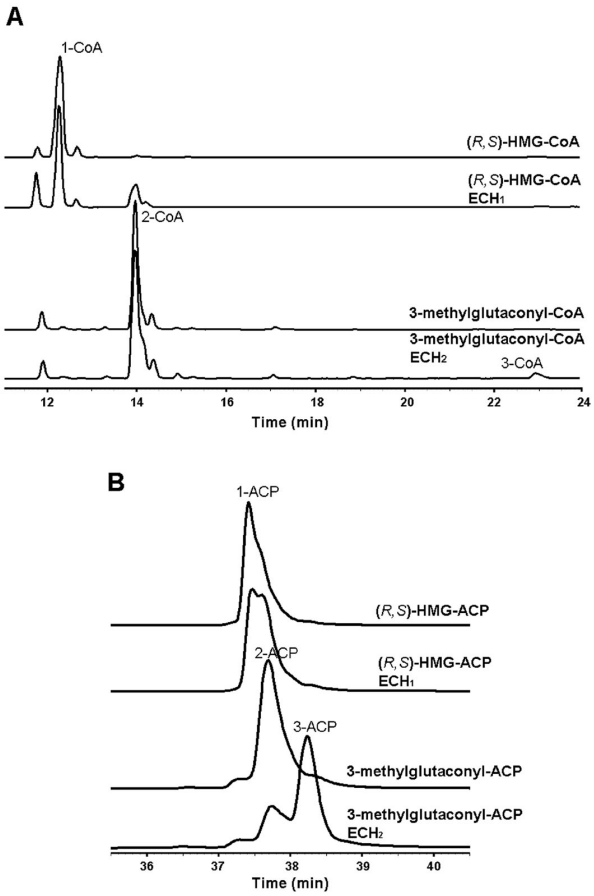



**Figure 3-6. Comparison of ACP and CoA substrate preference of $ECH_1$ and $ECH_2$. A**) CoA-dependent activities. HPLC traces were monitored at 254 nm. 50 μM (*R,S*)-HMG-CoA or 3-methylglutaconyl-CoA was treated with 2 μM $ECH_1$ or $ECH_2$. **B**) ACP-dependent activities. HPLC traces at 280 nm. 50 μM (*R,S*)-HMG-ACP or 3-methylglutaconyl-ACP was treated with 2 μM $ECH_1$ or $ECH_2$. *Peak 1*,(*R,S*)-HMG; *peak 2*, 3-methylglutaconyl; *peak 3*, 3-methylcrotonyl. The yields of CoA (**A**, *peak 2*) or ACP (**B**, *peak 2*) for the reactions of CoA (**A**, *peak 1*)/ACP (**B**, *peak 1*) treated with $ECH_1$ are similar (~10-15% conversion of substrate to product), but the yields of CoA (**A**, *peak 3*) and 3-ACP (**B**, *peak 3*) treated with $ECH_2$ are significantly different (~3% for the CoA substrate and >70% for the ACP substrate).

the oxyanion hole anchors the substrate in the active site and stabilizes the enolate anion, $His^{240}$ hydrogen bonds with the substrate carboxylate, and $Lys^{86}$ is the proton donor to the C-4 position of the decarboxylated substrate (Figure 3-7). The hydrophobic environment of the active site chamber enhances the reactivity of these groups.

Other than $Lys^{86}$ and $His^{240}$, $Tyr^{82}$ is the only polar side chain within the active site chamber. However, elimination of the phenolic hydroxyl by substitution with Phe had

only a two-fold effect on product formation, indicating that $Tyr^{82}$ does not play a major role in substrate hydrogen bonding or proton donation. Interestingly, $Tyr^{140}$ of MMCD was proposed to hydrogen bond with the carboxylate group of methylmalonyl CoA in a mechanism analogous to that outlined above (*26*). $Tyr^{82}$ of CurF $ECH_2$ and $Tyr^{140}$ of MMCD are not in analogous parts of the protein structures. Significantly, $Tyr^{82}$ is not conserved among sequences of decarboxylases in the crotonase superfamily (Figure 3-3C).

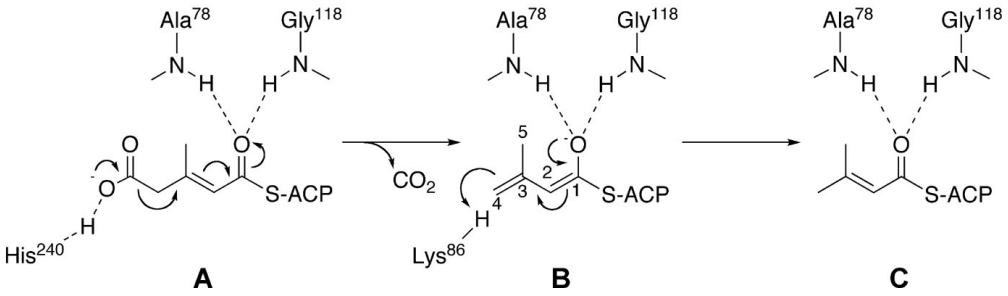



**Figure 3-7. Proposed CurF $ECH_2$ mechanism.** $His^{240}$ stabilizes the substrate carboxylate (**A**), and $Lys^{86}$ donates a proton to the decarboxylated intermediate (**B**) to form the product (**C**). The oxyanion hole is formed by the backbone amides of $Ala^{78}$ and $Gly^{118}$.

$His^{240}$ is conserved in all decarboxylases encoded by HCS cassettes and also in carboxymethylproline synthase (CarB) (Figure 3-3C). $His^{240}$ is in an excellent position to assist catalysis by hydrogen bonding with the carboxylate leaving group (Figure 3-7A). In liganded CarB, $His^{229}$ is in an identical position to $His^{240}$ of CurF $ECH_2$, and also was proposed to function as a general base in decarboxylation (*14*). The importance of $His^{240}$ in catalysis was demonstrated in the H240A and H240Q variants of CurF $ECH_2$, in which catalysis was severely impaired (Figure 3-5). $His^{240}$ of CurF $ECH_2$ is the only invariant polar residue within the active site chamber, and we propose that it acts as a general base here and in all other $ECH_2$-like decarboxylases of HCS cassettes (Figure 3-3C). An analogous His does not exist in MMCD. The equivalent helix to α10 of CurF $ECH_2$, which contains $His^{240}$, is further from the active site chamber in MMCD and forms extensive trimer contacts with the adjacent monomer.

After decarboxylation, proton donation to the substrate C-4 carbon to form

isopentenyl-ACP is necessary for enolate collapse to the $\Delta^2$ unsaturated product (Figure 3-7B). $Lys^{86}$ is the most likely proton donor in CurF $ECH_2$, given the relatively fixed position of the enolate anion in the oxyanion hole during substrate binding and the hydrophobic nature of the active site chamber. $Lys^{86}$ resides on helix α2, and has the necessary flexibility and reach to accommodate proton donation at the C-4 carbon of the substrate (Figure 3-4B). $His^{240}$ is not a candidate proton donor because it appears unable to reach the C-4 carbon of the decarboxylated intermediate, as required for subsequent product formation, and also unable to move closer to the C-4 carbon. Its backbone is held in place as part of the long helix α10 and its side chain is positioned by a hydrogen bond with the carbonyl of $Thr^{147}$ (Figure 3-4A) deep within the active site chamber. In contrast, $Lys^{86}$ is located on helix α2, the most mobile part of the CurF $ECH_2$ structure as seen in distinctly higher crystallographic temperature factors. This region of the crotonase fold is generally mobile and is completely disordered in some structures.

Proton addition to the C-4 carbon following decarboxylation should produce an α,β unsaturated product, whereas proton addition to C-2 should yield the β,γ product (Figure 3-7). Examples of both reaction routes exist, based on the structures of the PKS products. Thus, the mechanism of $ECH_2$ is a key determinant of the regiochemistry of the ultimate product of each pathway. The active site chambers of other $ECH_2$-like decarboxylases appear equally as hydrophobic as the CurF $ECH_2$ active site with the exception of the loop between helices α2 and α3 (residues 89-95 of CurF, Figure 3-3C). This region covers the active site (Figure 3-4B), but differs in length and sequence among the $ECH_2$ decarboxylases (Figure 3-3C). It is uncertain whether all residues aligned with CurF 89-95 constitute the α2-α3 loop in these decarboxylases, or whether some of them are part of a longer helix α3, as in CarB and MMCD (Figure 3-S5). Variability in the α2-α3 loop is expected among the $ECH_2$ decarboxylases. Each enzyme must possess a unique conformation to accommodate its substrate, often bulkier than the methyl group of the CurF $ECH_2$ substrate. In addition, a unique proton donor must be positioned to generate specifically an α,β or β,γ unsaturated product. Candidate proton donors include $Lys^{80}$, $Asp^{83}$ or $Asp^{84}$ of PksI, $His^{68}$, $Asp^{73}$ or $Asp^{77}$ of VirE, $Cys^{74}$ or $Asp^{75}$ of TaY, $Asp^{79}$ or $Asp^{82}$ of MupK, $Glu^{88}$, $Glu^{89}$, $Lys^{92}$ or $Asp^{95}$ of JamJ, and $Asp^{2949}$, $Lys^{2954}$, $Asp^{2962}$ or

$Glu^{2963}$ of PedI (Figure 3-3C).

The CurF $ECH_2$ domain is the first structurally characterized crotonase family member that acts preferentially upon ACP-linked substrates. Product formation *in vitro* with a CoA-linked substrate was significantly less robust than with an ACP-linked substrate (Figure 3-6), but the effect of each amino acid substitution was identical in ACP- and CoA-based assays (data not shown). Despite the strong preference for ACP-linked substrates by CurF $ECH_2$, structure alignment with other crotonase superfamily enzymes having CoA ligands (Figure 3-S5) revealed remarkably few structural changes in the CoA-binding region. Structural motifs that would recognize the adenine ring and overall backbone conformation around the adenosine portion of CoA are identical to other structurally characterized crotonase enzymes. One notable difference in CurF $ECH_2$ is the position of the $Tyr^{73}$ side chain. In the CoA-dependent enzymes, a basic side chain occupies this space and interacts with the 4-phosphate of pantothenic acid and the 5'-phosphate of the adenosine portion of CoA. In the $ECH_2$-like decarboxylases from HCS cassettes that operate with presumed ACP-linked substrates, $Tyr^{73}$ of CurF is conserved as a bulky phenylalanine or tyrosine residue (Figure 3-3C), whereas in the CoA-dependent enzymes the analogous position is always a smaller alanine, serine, or valine residue. Thus it appears that CoA is a poorer substrate of CurF $ECH_2$ in part because the $Tyr^{73}$ side chain has displaced a basic side chain ($Arg^{38}$ in CurF) from the CoA binding site. CurF $ECH_2$ containing an alanine substitution at the $Tyr^{73}$ site was not soluble and has frustrated efforts to directly test this hypothesis. The relatively small number of changes and persistent structural conservation implies a relatively recent evolution of protein function towards ACP-linked substrates.

An intriguing outcome of the CurF $ECH_2$ structure is the apparent symmetry mismatch between its trimeric structure and the fundamentally dimeric structure of several downstream domains in the CurF polypeptide (enoyl reductase, ketosynthase, dehydratase) (*6, 29*). The CurF $ECH_2$ domain possesses the crotonase "self-association" fold with the active site fully formed by each monomer and also demonstrates a capacity for trimer dissociation in solution (Figure 3-S4). Thus, the $ECH_2$ domain may be monomeric in the context of full-length CurF, similar to the bacterial fatty acid

β-oxidation multienzyme complex in which an N-terminal monomeric crotonase domain is fused to a dimeric dehydrogenase domain (*28*). Nevertheless, existence of the classical crotonase trimer in the isolated $ECH_2$ domain (Figure 3-3B) is strongly suggestive of a trimeric $ECH_2$ within full-length CurF. This could be accomplished in either of two ways. CurF could be a hexamer containing two $ECH_2$ trimers flexibly tethered to three dimers of the downstream domains. Alternatively, CurF could be a dimer in which $ECH_2$ forms a heterotrimer with $ECH_1$ (CurE). Determination of the oligomeric organization of CurF is an ongoing investigation.

In summary, the crystal structure of the N-terminal $ECH_2$ domain of CurF PKS-NRPS multifunctional protein from *Lyngbya majuscula* possesses a crotonase fold. Structural alignments identified $Ala^{78}$ and $Gly^{118}$ as forming the oxyanion hole, whereas substrate modeling identified $His^{240}$, $Lys^{86}$, and $Tyr^{82}$ as potential catalytic residues. Site directed mutagenesis and functional analysis using a coupled $ECH_1/ECH_2$ dehydration/decarboxylation assay demonstrated that $His^{240}$ and $Lys^{86}$ are critical for catalytic activity. Based on assay results, sequence conservation with decarboxylases from other HCS cassette-containing biosynthetic systems and comparisons with other structurally characterized crotonase decarboxylases, $His^{240}$ is proposed to be the general base involved in decarboxylation and $Lys^{86}$ is proposed to donate a proton to the C-4 position of product. These structural features enable CurF $ECH_2$ to specifically generate the α-β alkene regiochemistry in formation of the key isopentenyl-ACP product during curacin A biosynthesis.

## *3.4 MATERIAL AND METHODS*

**Chemical synthesis of 3-methyl-3-butenoyl-CoA.** 3-methyl-3-butenoic acid was prepared by oxidizing 3-methyl-3-buten-1-ol (Fluka) using $CrO_3/H_2SO_4$. Analytical data: $^1H$ NMR (300 MHz, $CD_3Cl$): δ =1.83 (s, 3H), 3.06 (s, 2H), 4.88 (s, 1H), 4.94 (s, 1H). 3-methyl-3-butenoyl-CoA was generated by the thioesterification reaction of coenzyme A (protocol from Professor Christopher T. Walsh). 6.5 μL 3-methyl-3-butenoic acid solution (100 mg/mL acid in THF), 200 μL (benzotriazol-1-yloxy) tripyrrolidinophosphonium hexafluorophosphate (PyBOP) suspension (66.5 mg/mL in THF) and 8.87 μL

N,N-diisopropylethylamine (DIPEA) were combined before adding 200 μL CoA solution (50 mg/mL). Reaction mixture was stirred for 3-5 hours at room temperature. The mixture was loaded onto a preparative $C_{18}$ column, and eluted with the gradient of 5-50% $CH_3OH/H_2O$ (10 mM $NH_4OAc$). Further purification was performed using a semipreparative $C_{18}$ column and the gradient of 5-40% $CH_3OH/H_2O$ (10 mM $NH_4OAc$). Analytical data: ESI-MS (-ve): expected $[M-H]^-$: 848.2, found: 848.3.

**Enzymatic synthesis of (*S*)-HMG-CoA.** (*S*)-HMG-CoA was enzymatically generated using HMG-CoA reductase as previously described (*24*). The 15 ml reaction mixture contained 60 mM Tris-HCl, pH 7.9, 60 mM (*R,S*)-mevalonate, 5 mM $NAD^+$, 1 mM CoA and 150 mM HMG-CoA reductase. 60 mM pyruvate and lactate dehydrogenase (200 units) were added to regenerate $NAD^+$. The reaction was stirred at room temperature for 3 hours and monitored at UV 340 nm. The precipitate and enzymes were removed from the solution using a 10 kDa cutoff membrane (Amicon Ultra, Millipore). The filtrated solution was applied to a preparative $C_{18}$ column using the elution gradient of 3-60% $CH_3OH/H_2O$ (10 mM $NH_4OAc$). The collected fraction containing (*S*)-HMG-CoA was evaporated and lyophilized. (*S*)-HMG-CoA was further purified by a semipreparative $C_{18}$ column using the elution gradient of 3-30% $CH_3OH/H_2O$ (10 mM $NH_4OAc$). Analytical data: ESI-MS (-ve): expected $[M-H]^-$: 910.2, found: 910.3.

**Bacterial strains, media and culture conditions.** *Escherichia coli* DH5α MCR (Invitrogen) was used for DNA propagation. *Escherichia coli* BL21 (DE3) (Invitrogen) transformed with the derivatives of pET24b and pET28b (Novagen) was used for protein overexpression in Luria-Bertani (LB) medium. Ampicillin (100 μg/mL), Kanamycin (50 μg/mL), Chloramphenicol (25 μg/mL), and Apramycin (50 μg/mL) were used for the corresponding plasmid construct resistance marker selection in *E. coli* cultures.

**Cloning, site-directed mutagenesis and protein overexpression.** CurB, CurE and CurF-$ECH_2$ (1-259 amino acid) genes were amplified from cosmid pLM54 described in our previous study (*6*) and inserted into pET24b and pET28b plasmids by the *Nde*I and *Xho*I restriction sites. The oligo pairs used for PCR amplification are the following: CurB ((**F**) 5'-ATT GGA GTT *CAT ATG* AGC AAA GAA CAA GTA-3', (**R**) 5'-CAA CAA

*CTC GAG* CAA TTT TGC TGC AAA TCA G-3'), CurE ((**F**) 5'-GAA CAC TAT *CAT ATG* TAT TAC AAA ACC ATA-3', (**R**) 5'-GAT TCA *CTC GAG* GTT CTG CCA TGG GTA ATA-3') and CurF-$ECH_2$ ((**F**) 5'-*CAT ATG* GCA GAA CTG AAT TTG AAT-3', (**R**) 5'-*CTC GAG* AAA CTC TTG CTG AAT ACG ACT-3'). The plasmid constructions were sequenced to confirm that they contained no errors. *E. coli* BL21 (DE3) was transformed by the expression plasmids, pET28b::CurE, pET24b::CurF-$ECH_2$ and pET28b::CurB, to express N-His-tagged CurE, C-His-tagged CurF-$ECH_2$, and N-His-tagged CurB. 0.8 liter cultures were inoculated with 8 ml of an overnight culture grown at 37°C. For CurB and CurF-$ECH_2$, the cultures were grown at 35°C to an OD (600 nm) = 0.5, and then cooled to 18°C prior to the addition of isopropyl-β-D-galactopyranoside (IPTG) (final concentration is 1 mM). The cultures were grown at 18°C for another 10-12 h. In order to increase the solubility of CurE, the pG-KJE8 chaperone system (TaKaRa) was used to coexpress the chaperones in *E. coli* according to the instruction manual. 2.0 mg/mL *L*-arabinose and 5 ng/mL tetracycline were added to the culture to induce the chaperone expression. The culture was grown at 30°C to an OD (600 nm) = 0.3-0.4, and cooled to 22°C before the addition of IPTG (0.5 mM). The culture was grown at 18°C for another 7-8 h prior to harvesting.

The plasmid pMCSG7::CurFd17 was generated by PCR amplification of coding sequence corresponding to residues 17-257 of CurF from the pML9 cosmid DNA (*6*) and inserted into the vector pMCSG7 (*30*). The plasmid pMCSG7::CurFd17 was transformed into BL21(*DE3*) and grown at 37°C in 2xYT media to an $OD_{600}$ of 0.6-0.8 in 2 L baffle flasks. The cultures were adjusted to 18°C, IPTG was added to final concentration of 0.4 mM, and allowed to grow 8-16 h with shaking. Cells were harvested by centrifugation and cell pellets were frozen immediately at -20°C. SeMet-labeled protein was produced in BL21 (DE3) using SeMet minimal media according to the protocol of Guerrero et al. (*31*). Site-directed mutagenesis was performed by QuikChange method (*Stratagene*) with the primers: **Y82F**: (**F**) 5'-CAT CAG GGG CTA GCA AAG AAT ***T***TT TAA TTA GAA AGA CTA GGG GTG AA AG-3', (**R**) 5'-CTT CAC CCC TAG TCT TTC TAA TTA AA***A*** ATT CTT TGC TAG CCC CTG ATG-3'; **K86A**: (**F**) 5'-GGG GCT AGC AAA GAA TAT TTA ATT AGA ***GC***G ACT AGG GGT GAA GTA GAA G-3', (**R**) 5'-CTT CTA CTT CAC CCC TAG TC***G C***TC TAA TTA AAT ATT CTT TGC TAG CCC C-3';

**K86Q**: (**F**) 5'-GGG GCT AGC AAA GAA TAT TTA ATT AGA ***C***AG ACT AGG GGT GAA GTA GAA G-3',(**R**) 5'-CTT CTA CTT CAC CCC TAG TCT ***G***TC TAA TTA AAT ATT CTT TGC TAG CCC C-3'; **D95A**: (**F**) 5'-GGG GTG AAG TAG AAG TTC TTG ***C***TT TAT CAG GGT TGA TAC-3', (**R**) 5'-GTA TCA ACC CTG ATA AA***G*** CAA GAA CTT CTA CTT CAC CCC-3'; **D95N**: (**F**) 5'-GGG GTG AAG TAG AAG TTC TT***A*** ATT TAT CAG GGT TGA TAC-3', (**R**) 5'-GTA TCA ACC CTG ATA AAT ***T***AA GAA CTT CTA CTT CAC CCC-3'; **H240A**: (**F**) 5'-GAA GCT ATC AAA AAA GAA CTA GAA ATA ***GC***C CAG GTA ACC TTT AAC CAA CC-3', (**R**) 5'-GGT TGG TTA AAG GTT ACC TGG ***GC***T ATT TCT AGT TCT TTT TTG ATA GCT TC-3'; **H240Q**: (**F**) 5'-GAA GCT ATC AAA AAA GAA CTA GAA ATA CA***A*** CAG GTA ACC TTT AAC CAA CC-3', (**R**) 5'-GGT TGG TTA AAG GTT ACC TG***T*** TGT ATT TCT AGT TCT TTT TTG ATA GCT TC-3'. All constructs were verified by DNA sequencing. CurA (2057-2146)-$ACP_{II}$ with a C-terminal $His_6$-tag was constructed by inserting synthetic DNA (a kind gift from Christopher Calderone and Christoper T. Walsh, Harvard Medical School) into pET29a using *Nde*I and *Xho*I restriction sites. Protein overexpression of CurA $ACP_{II}$ was performed as described for CurB ACP.

**Protein purification.** Protein purification was performed at 4°C. *E. coli* cells bearing an expression construct were harvested by centrifugation (4,000 x *g*, 15 min, 4°C), resuspended in the lysis buffer (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol) and disrupted by sonication. The cell debris was removed by centrifugation at 15,000 × *g* for 50 min. The supernatant was gently removed and loaded onto at the Ni-NTA agarose column pre-equilibrated with lysis buffer. The resin was washed successively with at least 5 column volumes of the washing buffer (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 20 mM imidazole) to remove nonspecifically bound contaminants. Bound proteins were eluted with the elution buffer (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 250 mM imidazole). The eluate was concentrated using Amicon Ultra centrifugal devices (Millipore) and immediately loaded onto the PD10 desalting column (GE Healthcare) pre-equilibrated with the storage buffer (50 mM PBS buffer, pH 7.5, 200 mM NaCl, 20% glycerol). The proteins were flash-frozen in liquid $N_2$ and stored at -80°C for further use. The purity of the protein was analyzed by SDS-PAGE and the protein concentrations were determined using the Bradford assay (Bio-Rad Protein assay,

Bio-Rad).

The protein purification for X-ray structural studies was performed in a similar way. The cell pellet from 1 L of culture was resuspended in 35 mL of buffer A (20 mM Tris pH 7.9, 500 mM NaCl, 10% glycerol, 20 mM imidazole). Cellytic Express (200-300 mg) was added prior to lysis by sonication (Sigma-Aldrich). Lysate was cleared by centrifugation at >38,000 x *g* for 45 minutes. Supernatant was filtered by 0.45 μm filter, loaded onto a 5 ml HisTrap (*GE Healthcare*) Ni-NTA resin column, and washed with 10 column volumes of buffer A. Protein was eluted with ~200 mM imidazole by a linear gradient of buffer B (20 mM Tris pH 7.9, 500 mM NaCl, 10% glycerol, 750 mM imidazole). For His-tag removal, fractions were pooled, buffer exchanged with buffer A containing 1 mM DTT, and incubated at 4°C for 24-72 h with 2% (w/w) His-tagged TEV protease. The reaction mixture was reloaded onto the HisTrap column and flow-through fractions were pooled, concentrated, and loaded onto a HiLoad 16/60 Superdex 75 (*GE Healthcare*) column equilibrated with 20 mM Tris pH 7.9, 500 mM NaCl, and 10% glycerol. Fractions were then combined, concentrated to 20 mg/ml and either flash frozen in liquid $N_2$ or stored at 4°C.

**FTMS analysis.** Multiply protonated (*R,S*)-HMG-CurB ACP with or without $ECH_1$ and/or $ECH_2$ treatment was generated by electrospray ionization (ESI) at 70 μL/h (Apollo ion source, Bruker Daltonics, Billerica, MA) of a solution containing 2.5 μM (*R,S*)-HMG-CurB (1:1 $CH_3CN$:$H_2O$ with 0.05% HCOOH and 0.05% $CF_3COOH$). To accurately determine the masses of HMG-CurB ACP and the species associated with the loss of $H_2O$ and ($H_2O$ + $CO_2$) from (*R,S*)-HMG-CurB ACP a calibration standard (G2421A, Agilent Technologies, Palo Alto, CA) was mixed with the sample (30-fold dilution of the standard) for internal calibration. All mass spectra were collected with an actively shielded 7 Tesla quadrupole-Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (APEX-Q, Bruker Daltonics). Ions produced by ESI were externally accumulated in a hexapole for 1 s, transferred via high voltage ion optics, and captured in an Infinity ICR cell (*32*) by gated trapping. This accumulation sequence was looped 32 times. The ESI capillary voltage was set to -3.8 kV. Nitrogen drying gas (200-250°C) was employed to assist desolvation of ESI droplets. The collision cell gas pressure (Ar)

was kept at $4.5 \times 10^{-6}$ mbar, which is slightly lower than the standard pressure used for this instrument ($6\text{-}7 \times 10^{-6}$ mbar). All data were acquired with XMASS software (version 6.1, Bruker Daltonics) in broadband mode from m/z = 200 to 2000 with 512k data points and summed over 16 or 32 scans. Mass spectra were analyzed with the MIDAS analysis software (*33*) Internal frequency-to-mass calibration was performed by Microsoft Excel with a two-term calibration equation (*34*) using the masses of two peaks (m/z = 622.02895 and 922.00979) resulting from the calibration standard. The theoretical and calculated masses are listed in table 3-S3. (Performed by Haichuan Liu)

**HPLC and mass spectrometry analysis.** The ECH activity assays using HMG-CoA substrate were performed by incubating 500 μM HMG-CoA and 2 μM $ECH_1$, $ECH_2$ or both in 30 mM Bis-Tris-HCl buffer, pH 6.5 at 37°C for 5 h. Before the LC-MS analysis, the reaction mixtures were filtered using a 10 kDa cutoff membrane (0.5 ml, Amicon Microcon, Millipore) to remove the enzymes. The filtered mixtures were analyzed by HPLC using a reverse-phase column chromatography (XBridge C18, 3.5 μm, 2.1x150 mm, Waters). The LC-MS was performed on the ThermoElectron Surveyor HPLC system equipped with the ThermoElectron Finnigan ESI-LTQ mass spectrometer. The isolation width for MS/MS-fragmentation was 1.5 and the normalized energy was set to 28%.

**Crystallization.** Crystals were grown in 24-72 h at 4°C by micro-seeding in hanging drops using the vapor diffusion method. Equal volumes were mixed of protein solution and mother liquor containing 1.3-1.8 M sodium malonate pH 7.0, 50 mM HEPES pH 6.8, and 0-25 mM 2-ethanamidoethyl 3-methylbut-2-enethioate. Crystals were harvested in loops and directly frozen by plunging into liquid $N_2$. Both trigonal (P321) and rhombohedral (R32) crystals grew in the above crystallization condition from the same seed stock. (Performed by Todd Geders)

**Data collection and structure determination.** Diffraction data were collected at 100 K on GM/CA-CAT beamlines 23ID-B and 23ID-D at the Advanced Photon Source (APS) in Argonne National Laboratory (Argonne, IL). Data were processed using the HKL2000 suite (*35*). Initial phasing was performed using a three-wavelength MAD dataset from a

single trigonal wild-type SeMet-labeled protein crystal. SOLVE was used to find the four selenium sites and for MAD phasing (‹m› = 0.47, score = 25.07) (*36*). RESOLVE was used for density modification (overall FOM = 0.69) and partial automated model building (*37, 38*). In both crystal forms, one $ECH_2$ subunit was present in the asymmetric unit. Modeling was completed manually using COOT (*39*). Refinement was performed using REFMAC5 of the CCP4 suite with TLS (*40-43*). The refined model was used as a probe structure for molecular replacement using PHASER with data from rhombohedral crystals (*43, 44*). (Performed by Todd Geders)

**Sequence and structure analysis.** Sequence alignments were performed by T_COFFEE (40). Figures and structure alignments were generated with PyMOL (*45*). Surface area calculations were calculated using AREAIMOL (*46*). Identification of structural neighbors was performed by a DALI search (*47*). (Performed by Todd Geders)

**Structure alignment and substrate modeling.** Cα atoms of the crotonase core domain of CurF $ECH_2$ (residues 18-222) were aligned with the analogous core region of liganded crotonase family members carboxymethylproline synthase ((*14*), PDB code 2A81, RMSD=2.3Å for 167 $C_\alpha$), methylmalonyl-CoA decarboxylase ((*26*), PDB code 1EF9, RMSD=1.7Å for 153 $C_\alpha$), 1,4-dihydroxy-2-naphthoyl-CoA synthase ((*48*), PDB code 1Q51, RMSD=1.6Å for 138 $C_\alpha$), rat enoyl-CoA hydratase ((*12*), PDB code 1DUB, RMSD=1.8Å for 156 $C_\alpha$), and 4-chlorobenzoyl-CoA dehalogenase (*13*), PDB code 1NZY, RMSD=2.0Å for 163 $C_\alpha$) (Figure 3-S5). Initial atomic coordinates and topology files for the substrate 3-methylglutaconyl moiety attached to 4-phosphopantetheinic acid were generated using the PRODRG2 server (*49*). Using the conformation of 4-phosphopantethenic acid bound to rat ECH as a guide (PDB code 1DUB), the substrate thioester oxygen was fixed at the position of a water molecule in the oxyanion hole of wild-type CurF $ECH_2$. Three water molecules overlapping the modeled substrate were removed. The model was refined by energy minimization using the program CNS (*50*) in 500 steps of conjugate gradient minimization with no experimental energy terms, no crystallographic symmetry restraints, and fixed main chain positions after addition of polar hydrogens. (Performed by Todd Geders)

**$ECH_1/ECH_2$ coupled enzymatic assay.** The activities of the CurF $ECH_2$ wild type and variants were measured in the $ECH_1$-$ECH_2$ coupled assay, as previously reported (*10*). In brief, 50 μM (*R,S*)-HMG-CurA-$ACP_{II}$ was incubated with 2 μM $ECH_1$ and $ECH_2$ (WT or variants) in 50 mM Tris-HCl (pH 7.5) at 37°C for 1 h. The reactions were terminated by 10% formic acid immediately before loading the reaction mixture on the Jupiter C4 (5μ, 300 A) reverse phase column (Phenomenex) and the proteins were eluted with $CH_3CN$ (0.05% HCOOH and 0.05% $CF_3COOH$)/$H_2O$ (0.05% HCOOH and 0.05%$CF_3COOH$). The chromatogram peaks were normalized by 32 Karat software (Beckman Coulter) and subjected to base-line subtraction before peak-area calculations.

**Preparation of 3-methylglutaconyl-CoA and ACP.** In order to separate the dehydration and decarboxylation steps catalyzed by $ECH_1$ and $ECH_2$, 3-methylglutaconyl-CoA was prepared by enzymatic dehydration of (*R,S*)-HMG-CoA. (*R,S*)-HMG-CoA was incubated with $ECH_1$ at 37°C for 5 h, and the dehydration product was isolated using XBridge Prep C18 column (Waters, 10 x 250 mm, 5 μm) under the similar HPLC conditions reported (*10*). The fractions were pooled and lyophilized. ~0.5 mg 3-methylglutaconyl-CoA was generated from 6 mg (*R,S*)-HMG-CoA. 3-methylglutaconyl-ACP was prepared with the Sfp protocol (*10*).

**$ECH_1$ and $ECH_2$ assays for ACP and CoA substrates.** (*R,S*)-HMG-CoA/ACP and 3-methylglutaconyl-CoA/ACP were employed to test the $ECH_1$ and $ECH_2$ activities. 2 μM $ECH_1$ or $ECH_2$ was incubated with 50 μM ACP or CoA substrates in 50 mM Tris-HCl (pH 7.5) at 37°C for 1 h. The CoA samples were analyzed by XBridge Prep C18 column (Waters, 4.6 x 250 mm, 5 μm) and eluted with MeOH/H2O (10 mM $CH_3COONH_4$), and ACP samples were analyzed by Jupiter C4 column and eluted with $CH_3CN$ (0.05% HCOOH and 0.05% $CF_3COOH$)/$H_2O$ (0.05% HCOOH and 0.05% $CF_3COOH$).

## 3.5 SUPPLEMENTARY FIGURES AND TABLES

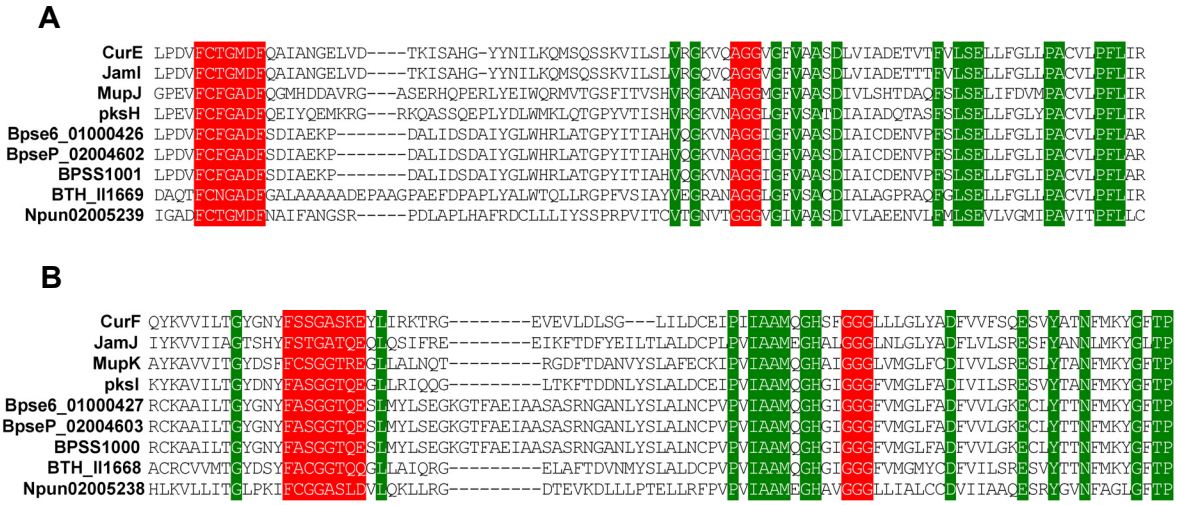



**Figure 3-S1. Partial sequence alignments of $ECH_1$ (A) and $ECH_2$ (B) from different microorganisms.** The two consensus sequences highlighted in red were reported to contribute to the formation of the oxyanion hole in the crotonase superfamily. The identical amino acids are highlighted in green. CurE, F: *Lyngbya majuscula*; JamI, J: *Lyngbya majuscula*; MupJ, K: *Pseudomonas fluorescens*; PksH, I: *Bacillus subtilis*; Bpse6_01000426, 01000427, *Burkholderia pseudomallei*; BpseP_02004602, 02004603: *Burkholderia pseudomallei*; BPSS1000, 1001: *Burkholderia pseudomallei*; BTH_II1668, 1669: *Burkholderia thailandensis*; Npun02005238, 02005239: *Nostoc punctiforme*.

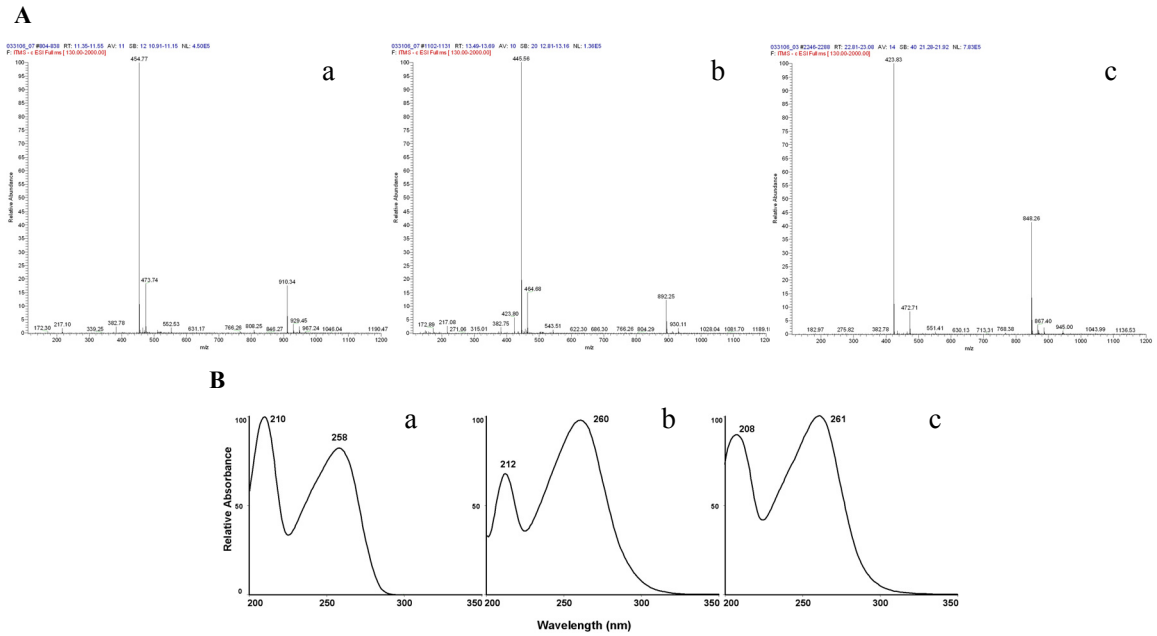



**Figure 3-S2. Mass spectra (A) and UV spectra (B) of CoA compounds. 5,** (HMG)-CoA (**a**), **6,** (3-methylglutaconyl)-CoA (**b**) and **7,** (3-methylcrotonyl)-CoA (**c**). The mass spectra and UV spectra were recorded at the retention times of the corresponding HPLC peaks after the background subtraction.

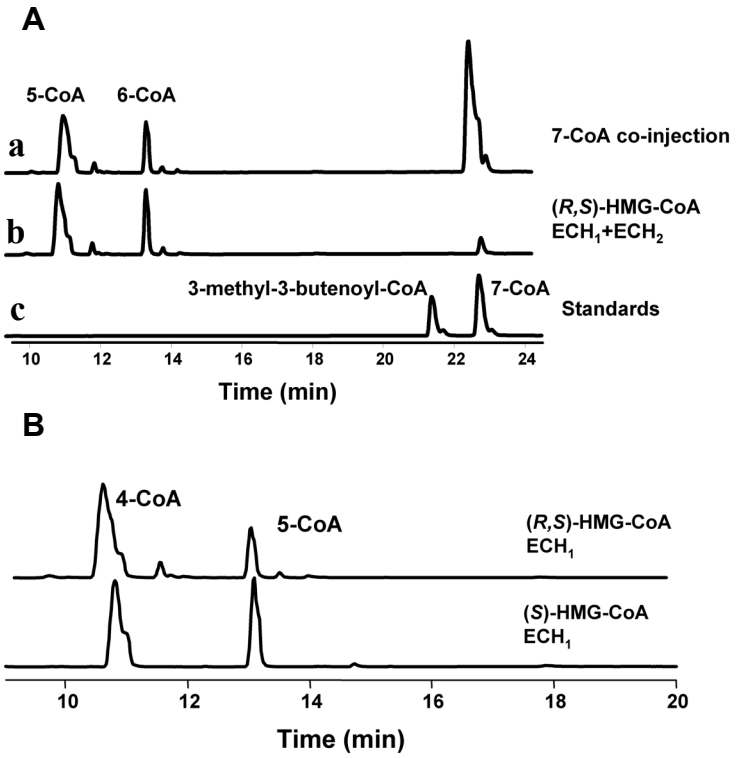



**Figure 3-S3. $ECH_1$ substrate specificity and $ECH_2$ regiochemical control. A**) HPLC chromatograms at 260 nm for $ECH_1$ and $ECH_1$ assays and co-injection with the standards. (**a**) Authentic standards: **7** (3-methylcrotonyl)-CoA and 3-methyl-3-butenoyl-CoA; (**b**). $ECH_1$/$ECH_2$ coupled assay using (*R,S*)-HMG-CoA; (**c**) the co-injection of (**b**) and **7**-CoA. **B**) HPLC chromatograms at 275 nm for $ECH_1$ assays using (*R,S*)-HMG-CoA and (*S*)-HMG-CoA. Reaction conditions: 500 μM HMG-CoA, 2 μM $ECH_1$, at 37°C for 1.5 hours. **5**-CoA: HMG-CoA, **6**-CoA: 3-methylglutaconyl-CoA.

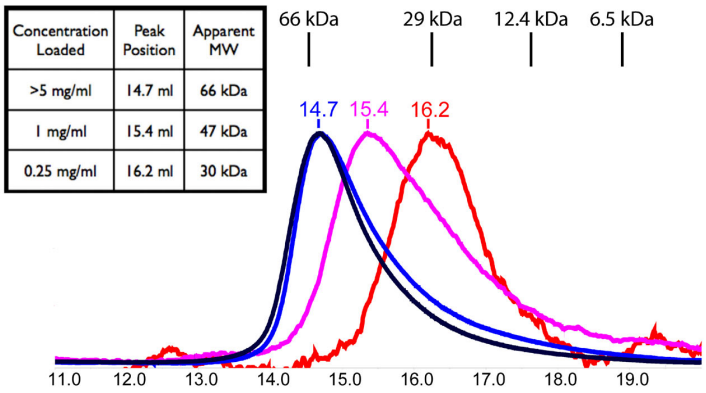



**Figure 3-S4. Analytical size-exclusion chromatography of the CurF $ECH_2$.** Change in apparent molecular weight is seen at loading concentrations below 5 mg/mL. Normalized A280 absorbance trace is shown for a 100 µL injection onto a 24 mL Superdex 200 10/300 GL column at the indicated protein concentration: black (10 mg/mL), blue (5 mg/mL), magenta (1 mg/mL), and red (0.25 mg/mL). Elution volumes for molecular weight standards are shown at the top. Similar elution patterns are seen with trimeric rat mitochondrial Δ3-Δ2-enoyl-CoA isomerase (*1*,*2*) and 4-hydroxycinnamoyl-CoA hydratase/lyase from Pseudomonas (*3*,*4*).

1. Palosaari, P. M., Kilponen, J. M., Sormunen, R. T., Hassinen, I. E., and Hiltunen, J. K. (1990) *J. Biol. Chem.* **265**(6), 3347-3353
2. Hubbard, P. A., Yu, W., Schulz, H., and Kim, J. J. (2005) *Prot. Sci.* **14**(6), 1545-1555
3. Mitra, A., Kitamura, Y., Gasson, M. J., Narbad, A., Parr, A. J., Payne, J., Rhodes, M. J., Sewter, C., and Walton, N. J. (1999) *Arch Biochem. Biophys.* **365**(1), 10-16
4. Leonard, P. M., Brzozowski, A. M., Lebedev, A., Marshall, C. M., Smith, D. J., Verma, C. S., Walton, N. J., and Grogan, G. (2006) *Acta Crystallogr. D* **62**(Pt 12), 1494-1501

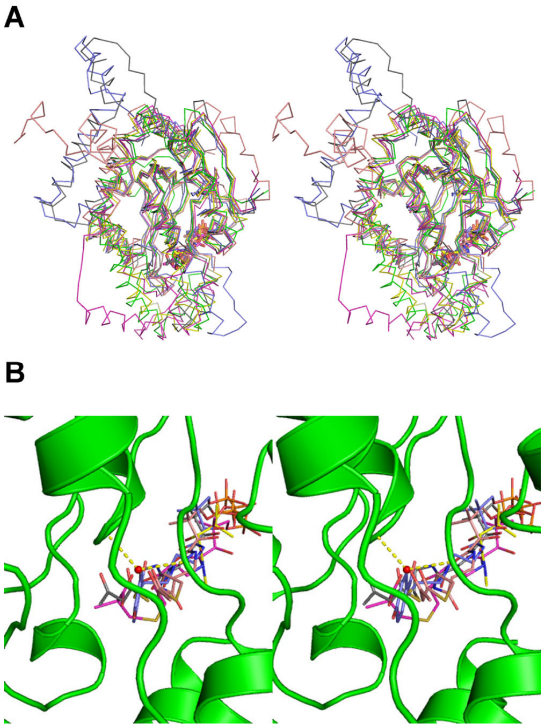



**Figure 3-S5. Superposition of CurF $ECH_2$ with other crotonase family members. A**, Stereo diagram of the superposition of wild-type CurF $ECH_2$ (green) with MMCD (magenta), CarB (yellow), MenB (pink), rat ECH (grey), 4-chlorobenzoyl CoA dehalogenase (blue). The 4-phosphopantetheine arms observed in the structures are shown in stick representation. The view is from the same orientation as Figure 3-4, with helix α3 present in front. **B**, Wild-type CurF $ECH_2$ is shown in cartoon representation (green). The water molecule in the oxyanion hole is shown as a red sphere with yellow dashed lines to backbone amides. The phosphopantetheine arm portion of the ligand from each structure in the superposition is shown with carbon coloring corresponding to the structure shown in A.

**Table 3-S1. Data Collection.**

| | Crystal Form I | | | | Crystal Form II | |
|---|---|---|---|---|---|---|
| Parameter | SeMet wt | Peak | Inflection | Remote | SeMet wt | Y82F |
| Space Group | P321 | P321 | P321 | P321 | R32 | R32 |
| Dimensions (Å) *a, c* | 105.4, 46.3 | 105.6, 46.1 | 105.7, 46.1 | 105.8, 46.2 | 106.2, 118.5 | 106.1, 119.5 |
| X-ray source | 23ID-B | 23ID-B | 23ID-B | 23ID-B | 23ID-D | 23ID-D |
| Wavelength λ (Å) | 1.0332 | 0.97942 | 0.97956 | 0.95373 | 1.0332 | 1.0332 |
| $d_{min}$ (Å) | 2.0 | 2.30 | 2.40 | 2.30 | 1.85 | 1.65 |
| Unique observations | 20,390 | 13,394 | 11,842 | 13,379 | 22,402 | 31,221 |
| $R_{merge}$(%)[a,b] | 6.3 (68) | 10.1 (66) | 10.2 (71) | 8.8 (67) | 7.2 (57) | 6.1 (61) |
| ‹I/σ›[a] | 36.2 (3.6) | 18.3 (2.9) | 18.4 (2.7) | 18.9 (2.3) | 27.7 (4.3) | 34.7 (3.7) |
| Completeness (%)[a] | 99.9 (100) | 100 (100) | 100 (100) | 100 (100) | 100 (100) | 99.9 (100) |
| Avg. redundancy[a] | 11.0 (10.4) | 5.9 (5.8) | 5.9 (5.8) | 5.9 (5.6) | 11.0 (10.3) | 10.9 (9.6) |

[a]Values in parenthesis are for outer shell
[b]$R_{merge}=\sum|I_i-\langle I\rangle|/\sum I_i$, where $I_i$ is the intensity of the *i*th observation and ‹I› is the mean intensity

**Table 3-S2. Refinement Statistics**

| Refinement statistics | Form I | Form II | |
|---|---|---|---|
| | SeMet wt | SeMet wt | Y82F |
| Space Group | P321 | R32 | R32 |
| Data range | 50-2.0 | 50-1.85 | 50-1.65 |
| $R/R_{free}$[a] | 0.202/0.250 | 0.169/0.206 | 0.167/0.204 |
| RMSD bond length (Å) | 0.011 | 0.012 | 0.012 |
| RMSD bond angle (Å) | 1.245 | 1.249 | 1.316 |
| Average Protein B-factor (Å$^2$) | 29.6 | 22.8 | 21.5 |
| Average Solvent B-factor (Å$^2$) | 44.6 | 41.0 | 40.0 |
| Wilson B (Å$^2$) | 33.2 | 25.6 | 23.6 |
| Ramachandran plot[c] | | | |
| Favored | 98.8% | 99.2% | 98.3% |
| Allowed | 1.2% | 0.8% | 1.7% |
| Disallowed | 0.0% | 0.0% | 0.0% |
| Protein atoms | 1899 | 1895 | 1894 |
| Water molecules | 145 | 184 | 307 |
| PDB code | 2Q2X | 2Q34 | 2Q35 |

[a]$R = \sum|F_O-|F_C||/\sum|F_O|$ where $F_O$ is the observed structure factor and $F_C$ is the calculated structure factor used in the refinement
[b]$R_{free} = \sum|F_O-|F_C||/\sum|F_O|$ where $F_O$ is the observed structure factor and $F_C$ is the calculated structure factor from 5% of reflections not used in the refinement
[c]From output of MolProbity.

**A**

| | Calculated masses | Experimental average masses | Error (ppm) |
|---|---|---|---|
| HMG-ACP(CurB) | 11325.84 | 11325.82 ± 0.16 | 1.09 |
| $H_2O$-loss peaks | 11307.82 | 11307.82 ± 0.14 | 0.40 |
| ($H_2O$ + $CO_2$)-loss peaks | 11263.84 | 11264.84 ± 0.25 | 0.25 |

**B**

| | Calculated masses | Experimental average masses | Error (ppm) |
|---|---|---|---|
| HMG-$ACP_{II}$ | 11785.91 | 11785.89 ± 0.06 | 1.21 |
| $H_2O$-loss peaks | 11767.90 | 11767.97 ± 0.03 | 5.70 |
| ($H_2O$ + $CO_2$)-loss peaks | 11723.91 | 11723.94 ± 0.05 | 2.50 |

**Table 3-S3.** Mass calibration results for **A**) HMG-ACP (CurB), **B**) HMG-$ACP_{II}$ and their associated peaks corresponding to the loss of $H_2O$ and ($H_2O$ + $CO_2$). The masses of the most abundant isotopes of the $8^+$-$15^+$ charge states were used for mass calculation.

## 3.6 REFERENCES


1. A. M. Hill, *Nat. Prod. Rep.* **23**, 256 (2006).
2. J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **18**, 380 (2001).
3. C. T. Walsh, *Science* **303**, 1805 (2004).
4. J. L. Fortman, D. H. Sherman, *Chembiochem* **6**, 960 (2005).
5. P. Verdier-Pinard *et al.*, *Mol. Pharmacol.* **53**, 62 (1998).
6. Z. X. Chang *et al.*, *J. Nat. Prod.* **67**, 1356 (2004).
7. D. J. Edwards *et al.*, *Chem. Biol.* **11**, 817 (2004).
8. F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. E. O'Connor, C. T. Walsh, *Nature* **436**, 1191 (2005).
9. A. K. El-Sayed *et al.*, *Chem. Biol.* **10**, 419 (2003).
10. L. C. Gu *et al.*, *J. Am. Chem. Soc.***128**, 9014 (2006).
11. J. A. Gerlt, P. C. Babbitt, *Annu. Rev. Biochem.* **70**, 209 (2001).
12. C. K. Engel, M. Mathieu, J. P. Zeelen, J. K. Hiltunen, R. K. Wierenga, *EMBO J* **15**, 5135 (1996).
13. M. M. Benning *et al.*, *Biochemistry* **35**, 8103 (1996).
14. M. C. Sleeman, J. L. Sorensen, E. T. Batchelar, M. A. McDonough, C. J. Schofield, *J. Biol. Chem.* **280**, 34956 (2005).
15. P. A. Hubbard, W. F. Yu, H. Schulz, J. J. P. Kim, *Prot. Sci.* **14**, 1545 (2005).
16. A. M. Mursula, D. M. F. van Aalten, J. K. Hiltunen, R. K. Wierenga, *J. Mol. Biol.* **309**, 845 (2001).
17. S. T. Partanen *et al.*, *J. Mol. Biol.* **342**, 1197 (2004).
18. P. M. Leonard *et al.*, *Acta Crystallogr D* **62**, 1494 (2006).
19. J. L. Whittingham, J. P. Turkenburg, C. S. Verma, M. A. Walsh, G. Grogan, *J. Biol. Chem.* **278**, 1744 (2003).
20. H. M. Holden, M. M. Benning, T. Haller, J. A. Gerlt, *Acc. Chem. Res.* **34**, 145 (2001).
21. T. W. Geders *et al.*, *J. Biol. Chem.* **282**, 35954 (2007).
22. R. H. Lambalot *et al.*, *Chem. Biol.* **3**, 923 (1996).
23. F. Lynen, S. Ochoa, *Biochim. Biophys. Acta* **12**, 299 (1953).
24. K. M. Bischoff, V. W. Rodwell, *Biochem. Med. Metab. Biol.* **48**, 149 (1992).
25. B. J. Wong, J. A. Gerlt, *Biochemistry* **43**, 4646 (2004).
26. M. M. Benning, T. Haller, J. A. Gerlt, H. M. Holden, *Biochemistry* **39**, 4630 (2000).
27. B. Gerratana, S. O. Arnett, A. Stapon, C. A. Townsend, *Biochemistry* **43**, 15936 (2004).

28. M. Ishikawa, D. Tsuchiya, T. Oyama, Y. Tsunaka, K. Morikawa, *EMBO J.* **23**, 2745 (2004).
29. D. H. Sherman, J. L. Smith, *ACS Chem. Biol.* **1**, 505 (2006).
30. L. Stols *et al.*, *Prot. Express. Purif.* **25**, 8 (2002).
31. S. A. Guerrero, H. J. Hecht, B. Hofmann, H. Biebl, M. Singh, *Appl. Microbiol. Biotechnol.* **56**, 718 (2001).
32. P. Caravatti, M. Allemann, *Org. Mass Spectrom.* **26**, 514 (1991).
33. M. W. Senko, J. D. Canterbury, S. H. Guan, A. G. Marshall, *Rapid Commun. Mass Spectrom.* **10**, 1839 (1996).
34. E. B. Ledford, D. L. Rempel, M. L. Gross, *Anal. Chem.* **56**, 2744 (1984).
35. Z. Otwinowski, W. Minor, in *Macromol. Crystallogr., Pt A*. (1997), vol. 276, pp. 307-326.
36. T. C. Terwilliger, J. Berendzen, *Acta Crystallogr D* **55**, 849 (1999).
37. T. C. Terwilliger, *Acta Crystallogr D* **59**, 38 (2003).
38. T. C. Terwilliger, *Acta Crystallogr D* **56**, 965 (2000).
39. P. Emsley, K. Cowtan, *Acta Crystallogr D* **60**, 2126 (2004).
40. G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr D* **53**, 240 (1997).
41. S. Bailey, *Acta Crystallogr D* **50**, 760 (1994).
42. J. Painter, E. A. Merritt, *Acta Crystallogr D* **62**, 439 (2006).
43. L. C. Storoni, A. J. McCoy, R. J. Read, *Acta Crystallogr D* **60**, 432 (2004).
44. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr D* **61**, 458 (2005).
45. W. L. DeLano, *The PyMOL Molecular Graphics System*, http://www.pymol.org (2002).
46. B. Lee, F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
47. L. Holm, C. Sander, *Prot-Struct. Funct. Bioinform.* **33**, 88 (1998).
48. J. J. Truglio *et al.*, *J. Biol. Chem.* **278**, 42352 (2003).
49. A. W. Schuttelkopf, D. M. F. van Aalten, *Acta Crystallogr D* **60**, 1355 (2004).
50. A. T. Brunger *et al.*, *Acta Crystallogr D* **54**, 905 (1998).

**Notes:**

Liangcai Gu and David Sherman designed and performed the experiments;

Todd Geders and Janet Smith solved the Cur $ECH_2$ structure and designed the site-directed mutagenesis;

Haichuan Liu and Kristina Håkansson recorded and analyzed the FT-ICR mass spectra.

# CHAPTER 4
# HALOGENATION, CYCLOPROPANATION AND POLYKETIDE DIVERSIFICATION

## *4.1 SUMMARY*

The chemical diversity of natural products is fueled by the emergence and ongoing evolution of biosynthetic pathways in secondary metabolism (*1-5*). However, the co-evolution of enzymes as functional assemblies for metabolic diversification is not well understood, especially at the biochemical level. Here, two parallel enzyme assemblies with an extraordinarily high sequence identity, including a halogenase domain (Hal), a 3-hydroxy-3-methylglutaryl (HMG) enzyme cassette and an enoyl reductase domain (ER) in the curacin A (Cur) and jamaicamide (Jam) pathways catalyze the formation of β-branched cyclopropane and vinyl chloride moieties respectively, in their final products. Bioinformatic analysis indicated that their corresponding genes were inserted into modular polyketide synthases (PKSs) via acyltransferase (AT) domain replacement. The Hal from CurA, and dehydratases ($ECH_1$s), decarboxylases ($ECH_2$s) and ERs from both Cur and Jam were biochemically assessed to determine the mechanism of cyclopropane and vinyl chloride group formation. In the Cur pathway, $ECH_1$, $ECH_2$ and ER were presumed to catalyze successive dehydration, decarboxylation, and α,β C=C (enoyl thioester) saturation from an (*S*)-HMG substrate tethered to a tandem acyl carrier protein tridomain ($ACP_3$), to generate isovaleryl-$ACP_3$. Unexpectedly, this polyketide β-branching scheme was modified by introduction of a γ-chlorination step on (*S*)-HMG group mediated by Cur Hal, a non-haem $Fe^{II}$, α-ketoglutarate (α-KG)-dependent halogenase (*6*). CurF ER was found to specifically catalyze an unprecedented cyclopropanation reaction of the chlorinated substrate. In contrast, regiochemical control by Jam $ECH_2$ selectively generates the β,γ C=C (enoyl thioester) of the

3-methylglutaconyl-$ACP_3$ decarboxylation product, thus forming a vinyl chloride, and rendering the Jam ER redundant. Moreover, site-directed mutagenesis experiments demonstrated that $Tyr^{82}$ of Cur $ECH_2$ is crucial for regiochemical control. Thus, the combination of chlorination and polyketide β-branching, coupled with mechanistic diversification of $ECH_2$ and ER leads to formation of the cyclopropane and vinyl chloride functional groups. These results show how the unique interplay of modular PKS domain integration and β-branching enzyme evolution generates functional group diversity in secondary metabolites.

## *4.2 INTRODUCTION*

The tremendous biosynthetic capability of nature is well exemplified by structurally diverse secondary metabolites that help their hosts, typically microorganisms and plants, to survive and thrive in particular environmental niches by mediating a broad range of ecological and physiological interactions (*1*-*5*). The biosynthesis of secondary metabolites is "diversity-oriented" (*1, 5, 7*), targeting the variable environment by producing a vast array of complex chemical structures (*8*). This enormous productivity is largely fueled by the rapid evolution of biosynthetic genes and functional alteration of the corresponding enzymes (*3, 4*). As such, the evolutionary history of metabolic gene assemblies informs the origin of their biosynthetic diversity. However, tracing the ancestral forms of multiple genes as a functional collective is elusive, especially when they are dispersed in the genome. Biosynthetic genes from microbial hosts are usually clustered in their genomes, and as such are ideal for evolutionary and functional studies (*4*). To date, our understanding of the evolution of biosynthetic systems is largely limited to the genetic level. Recently, an expansion of sequencing information is providing rapid access to natural product biosynthetic enzymes, which is enabling efforts to trace the origin of metabolic diversity and to exploit biochemical diversity.

Modular PKSs originated from fatty acid synthases (FASs) and serve as a paradigm for secondary metabolic systems that are successful in natural selection to expand chemical diversity. These giant biosynthetic machines are modularized assembly-lines catalyzing highly programmed metabolic pathways. The chain extension modules

minimally comprise three essential domains—ketosynthase (KS), AT, ACP—plus auxiliary processing domains (e.g., ketoreductase (KR), dehydratase (DH), and ER) for β-keto group modification. Chemical diversity of the final products tends to be generated by introduction of variation in chain initiation, extension and termination steps (*9-11*). Similar to other enzymes in secondary metabolism, PKS catalytic domains show significant substrate tolerance to expand pathway chemical diversity (*9*). Moreover, PKS systems are prone to form hybrids with other biosynthetic elements, such as nonribosomal-peptide synthetase (NRPS) modules, and recently reported HMG enzyme cassettes (*12, 13*). The metamorphic properties shown in these composite systems drive metabolic diversification, but, remain poorly understood.

The curacin and jamaicamide marine cyanobacterial metabolites from *Lyngbya majuscula* are mixed-polyketide nonribosomal-peptide natural products with potent anticancer and sodium channel blocking activities, respectively (*14, 15*). The parallel components of the Cur and Jam biosynthetic pathways (Figure 4-1a) provide an unusual opportunity to investigate the biosynthetic origin of chemical diversity, in the form of cyclopropane ring formation for curacin and vinyl chloride formation for jamaicamide (*14, 16*). Studies on the function and selectivity of these highly parallel biosynthetic systems form the subject of this report.

## 4.3 RESULTS

### 4.3.1 Two Highly Similar Enzyme Assemblies

The two parallel, highly conserved Cur and Jam pathway enzyme assemblies are incorporated into the early PKS modules, and are predicted to catalyze β-branching type modification of the growing chain elongation intermediate (*14, 16*). These unusual embedded domains and discrete enzymes span from CurA to CurF and from JamE to JamJ, and are grouped into three subsets (Figure 4-1a): (1) Hals embedded in CurA and JamE; (2) HMG enzyme cassettes containing $ACP_3$ tridomains (tandem $ACP_I$, $ACP_{II}$ and $ACP_{III}$) embedded in CurA and JamE, discrete CurB and JamF $ACP_{IV}$s, CurC and JamG KSs, CurD and JamH HMG-CoA synthase-like enzymes (HCSs), CurE and JamI $ECH_1$s, $ECH_2$s embedded in CurF and JamJ; and (3) ERs embedded in CurF and JamJ

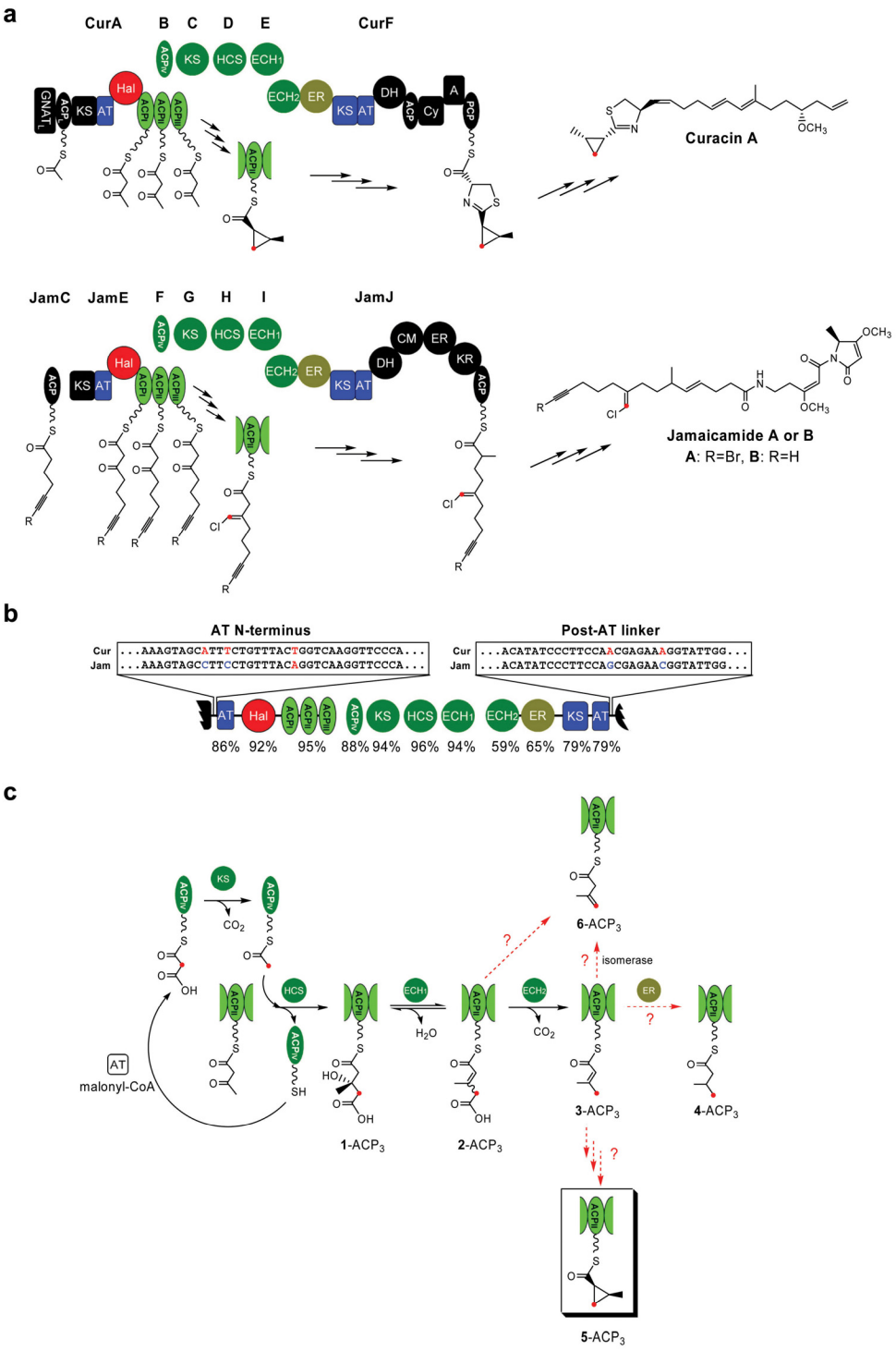



**Figure 4-1. Comparison of enzyme assemblies in the Cur and Jam pathways**. **a**, Formation of cyclopropane and vinyl chloride functional groups. **b**, Comparative sequence identities of the HMG-type enzyme assemblies. The aligned DNA sequences are located at the boundaries of these two highly related regions. **c**, Formation of **3**-$ACP_3$ in the Cur pathway, and hypothesized reactions for **4**-$ACP_3$, **5**-$ACP_3$ and **6**-$ACP_3$. The β-branching carbon atoms are highlighted in red.

enzymes revealed that the sequence identities of the Hals, $ACP_3$s, $ACP_{IV}$s, KSs, HCSs and $ECH_1$s are extraordinarily high (~90%), whereas the $ECH_2$s and ERs show significantly lower sequence identity (~60%) (Figure 4-1b).

Cur and Jam Hals show low homology (less than 20% sequence identity) to the α-KG-dependent non-haem enzyme superfamily, the members of which usually catalyze oxidation reactions (*17*). Some recently characterized halogenases belong to this superfamily (*18-20*) and are able to catalyze halogenation of unactivated carbon atoms (*19-23*) through a non-haem $Fe^{IV}$=O intermediate (*24, 25*). The presence of a β-branched vinyl chloride in the jamaicamide structure suggested that Jam Hal was responsible for the key halogenation step. By analogy, it suggested that transient halogenation might be crucial for curacin cyclopropane ring formation (*6, 19*).

HMG enzyme cassettes have been demonstrated to catalyze polyketide β-branching to generate a pendant methyl or ethyl group from a polyketide β-carbonyl (*12, 13, 26*). As shown for curacin A (Figure 4-1c), the AT loads a malonyl group from malonyl-CoA to CurB $ACP_{IV}$, and the KS catalyzes subsequent decarboxylation of malonyl-$ACP_{IV}$ to acetyl-$ACP_{IV}$. HCS then catalyzes condensation of C-2 from acetyl-$ACP_{IV}$ and acetoacetyl-$ACP_3$, to form (*S*)-HMG-$ACP_3$ (**1**-$ACP_3$). As we have shown previously, $ECH_1$ catalyzes dehydration of **1**-$ACP_3$ to 3-methylglutaconyl-$ACP_3$ (**2**-$ACP_3$), followed by $ECH_2$ mediated decarboxylation to generate 3-methylcrotonyl-$ACP_3$ (**3**-$ACP_3$) (*12*). Notably, chain elongation intermediates on $ACP_3$ are presumed to undergo multiple modification steps before transfer as substrates to downstream enzymes. Presumably, the tridomain $ACP_3$ facilitates metabolic efficiency, as suggested in other FAS and PKS systems bearing tandem ACPs (*27, 28*).

Sequence analysis of Cur and Jam ERs (Figure 4-1b) suggested that each functions with the adjacent HMG enzyme cassette based on a presumed ability to catalyze reductive conversion of α,β C=C (enoyl thioester) in **3**-$ACP_3$ to isovaleryl-$ACP_3$ (**4**-$ACP_3$). They show ~50% sequence identity to the other PKS ERs in Cur and Jam pathways, and belong to the acyl-CoA reductase family that catalyzes NADPH-dependent reduction of α,β C=C (enoyl thioester) in acyl-CoAs or acyl-ACPs

(*29*). These two ERs are located upstream of CurF and JamJ KS, an unusual location as ERs typically reside between AT and ACP domains in PKS modules. A similar location was reported for TaO ER in the myxovirescin A biosynthetic pathway, also predicted to catalyze reduction of the α,β C=C (enoyl thioester) of the corresponding $ECH_2$ decarboxylation product (*30*). Thus, the Cur and Jam HMG cassette enzymes including $ACP_3$, $ACP_{IV}$, KS, HCS, $ECH_1$, $ECH_2$ and ER comprise an integrated functional assembly for polyketide chain β-branching. Some HMG-type cassettes in other pathways lack one or more enzymes, and are occasionally found to have a dispersed architecture (*31-33*).

Our previous work on Cur $ECH_1$ and $ECH_2$ revealed enzymatic formation of **3**-$ACP_3$ (*12, 34*), a presumed precursor for (1*R*, 2*S*)-2-methylcyclopropane-1-carboxyl-$ACP_3$ (**5**-$ACP_3$) (Figure 4-1c), but the precise steps leading to cyclopropane ring formation remained to be established. This initial study raised a number of additional questions: (1) What is the timing and regiochemistry of the presumed Hal reaction, and what is the specific substrate? (2) Is Cur ER involved in reduction of **3**-$ACP_3$? (3) How is the pendant vinyl chloride formed in the Jam pathway, and what is the key control point for the introduction of this functional group? As previously proposed, could 3-methyl-3-butenoyl-$ACP_3$ (**6**-$ACP_3$) be generated from **3**-$ACP_3$ isomerization (*6*) or by differential regiochemical control of double bond formation during $ECH_2$ decarboxylation (*34*)?

### *4.3.2 AT Replacement-Mediated PKS Hybridization*

Bioinformatic analyses of *cur* and *jam* cluster sequences suggested that the DNA fragment that encompasses the parallel Cur and Jam AT-Hal-$ACP_I$-$ACP_{II}$-$ACP_{III}$-$ACP_{IV}$-KS-HCS-$ECH_1$-$ECH_2$-ER-KS-AT gene assembly might have been introduced into the polyketide pathway by AT domain replacement (Figure 4-1b). Based on the DNA and amino acid alignments of CurA—CurF and JamE—JamJ regions, we found that the highly similar regions, with clear demarcation in their DNA and protein sequences, extend from the N-termini of the ATs in CurA and JamE, through the C-terminal "post-AT linkers" (*35*) of the ATs in CurF and JamJ

(Figure 4-1b). Moreover, the two sites that delimit these regions are conserved for the ATs from the sequenced *cur* and *jam* gene clusters isolated from unique strains of *L. majuscula* (Figure 4-S1). Recent bioinformatic studies indicate that these highly similar sequences could promote AT domain replacement by homologous recombination (*36, 37*). Thus, this gene assembly could have been introduced by a "di-AT replacement", which could represent a prevalent strategy for pathway expansion or contraction in modular PKS systems.

To further probe our hypothesis, we constructed phylogenetic trees for KSs, ATs and DHs of the sequenced pathways from *L. majuscula*. The CurF and JamJ KSs within the proposed di-AT replacement are most closely related, but in contrast the CurA and JamE KSs outside this region are distant (Figure 4-S2). CurA and JamE ATs as well as CurF and JamJ ATs reveal extraordinarily high similarity (86% and 79% sequence identities, respectively). As expected, the CurF and JamJ DHs, which are outside the di-AT replacement region, show a relatively distant relationship, with the JamJ DH being most similar to DHs from the putative carmabin pathway (~70% sequence identity). Overall, this phylogenetic analysis is consistent with the di-AT replacement hypothesis.

### *4.3.3 HMG β-Branching with ER Saturation*

To assess CurF ER function as a putative reductase, the embedded domain was excised and cloned as an N-terminal GST-tagged fusion protein. We also overexpressed and purified the CurA $ACP_3$ tridomain and each excised single domain ($ACP_I$, $ACP_{II}$ and $ACP_{III}$) as apo proteins (Figure 4-S3). **1**-ACPs were generated as previously described[12], and substrate loading was examined by HPLC (Figure 4-S4). The $ACP_I$, $ACP_{II}$ and $ACP_{III}$ have nearly identical amino acid sequences, and each was efficiently loaded with the HMG substrate. Thus, for convenience we chose excised CurA $ACP_{II}$, as well as $ACP_3$, for subsequent enzyme assays.

Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) and infrared multiphoton dissociation (IRMPD) methods were applied to detect mass changes to the HMG substrate covalently linked to the ACP phosphopantetheine (PPant) arm (*38*). Cur ER function was assessed by coupling it with the Cur $ECH_1$ and $ECH_2$ reactions. As

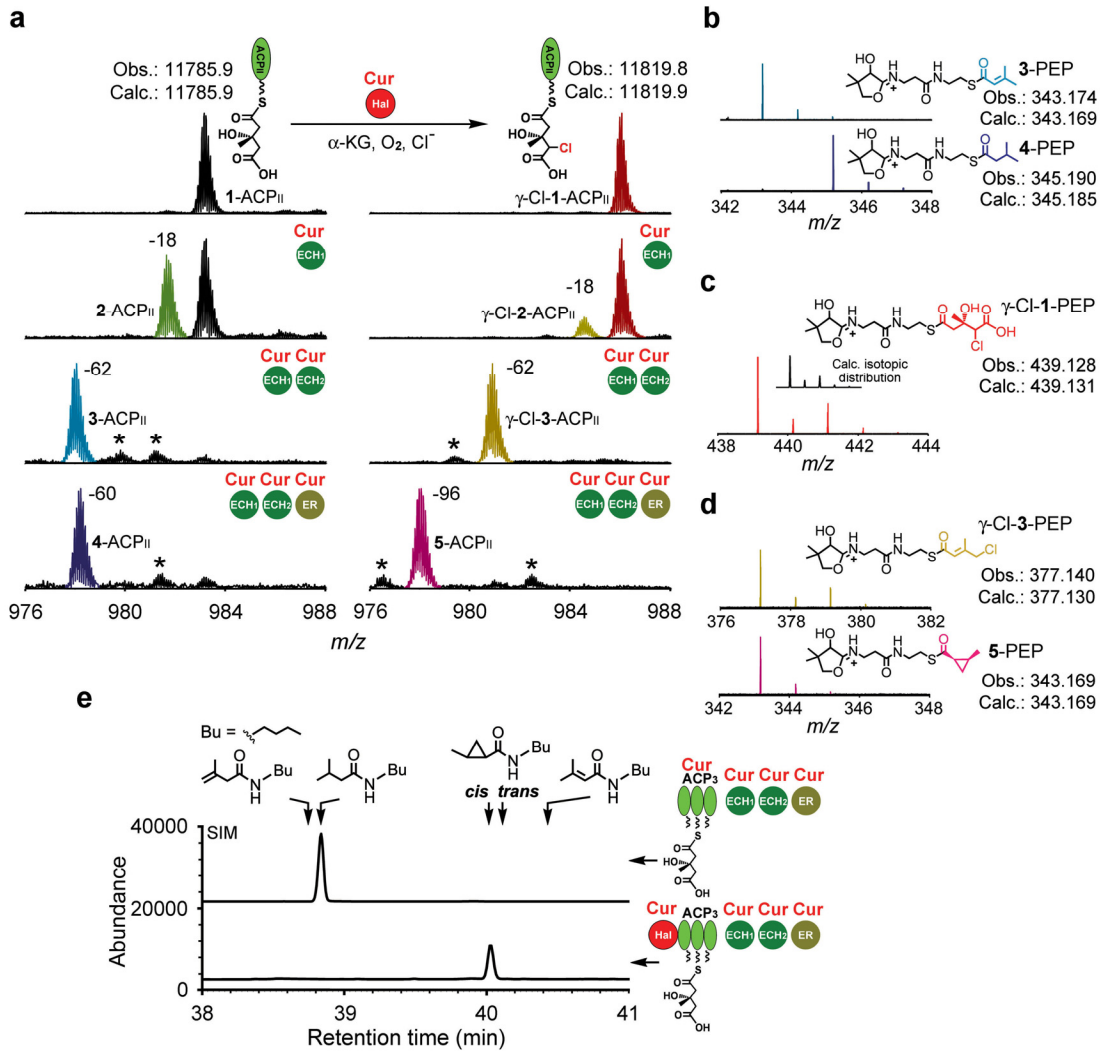



**Figure 4-2. Halogenation and cyclopropanation in the Cur pathway**. **a**, Partial FTICR mass spectra (12+ change state of $ACP_{II}$) for Cur Hal, $ECH_1$, $ECH_2$ and ER reactions. **1**-$ACP_{II}$ was incubated with Hal for 2 h to generate the γ-Cl-**1**-$ACP_{II}$ substrate. Reactions were incubated at 30°C for 2 h for the **1**-$ACP_{II}$ substrate and 30 min for the γ-Cl-**1**-$ACP_{II}$ substrate. Asterisks denote unidentified species. **b**, IRMPD spectra for the PPant ejection products (PEPs) of **3**-$ACP_{II}$ and **4**-$ACP_{II}$. **c**, IRMPD spectrum for the PEP of γ-Cl-**1**-$ACP_{II}$ and its theoretical isotopic distribution. **d**, IRMPD spectra for the PEPs of γ-Cl-**3**-$ACP_{II}$ and **5**-$ACP_{II}$. **e**, GC-MS analysis of the enzyme products after butylamine cleavage, and comparison with authentic standards. For optimal sensitivity, the chromatograms were recorded at selective ion mode (SIM) by monitoring 55, 57, 83, 115, 155 and 157 atomic mass unit (amu). Retention times of the products were confirmed by co-injection with authentic standards.

reported, Cur $ECH_1$ catalyzed the reversible dehydration of **1**-$ACP_{II}$ to generate **2**-$ACP_{II}$, and Cur $ECH_2$ catalyzed decarboxylation of **2**-$ACP_{II}$ to generate **3**-$ACP_{II}$, corresponding

to 18- and 62-Dalton mass losses from **1**-$ACP_{II}$, respectively (Figure 4-2a). Cur $ECH_1$ shows substrate preference for (*S*)-HMG-$ACP_{II}$ over (*R*)-HMG-$ACP_{II}$ (Figure 4-S5), which is consistent with our previous results using the CoA-linked substrates (*12*). With Cur ER and NADPH, a 2-Dalton mass addition was observed for **3**-$ACP_{II}$ (Figure 4-2a and 2b), corresponding to saturation of the α,β enoyl thioester to generate **4**-$ACP_{II}$. No saturation product was detected by using Cur ER and NADH as cofactor.

To confirm the structure of the Cur ER reaction product tethered to $ACP_3$, we cleaved it from the PPant arm with butylamine to generate the corresponding butylamide derivative. Gas chromatography (GC)-MS analysis was performed and the readily separable isomers were compared and correlated with authentic standards (*39*). We used **1**-$ACP_{II}$ and **1**-$ACP_3$ as substrates for Cur $ECH_1$, $ECH_2$ and ER reactions. According to the retention time and mass spectrum of the product, its structure was confirmed as **4** (Figure 4-2e). These experiments demonstrated that a reductive step catalyzed by Cur ER is included in the polyketide β-branching series of reactions and that ER is a component of the HMG enzyme cassette. Based on this analysis, **4**-$ACP_{II}$ had been assumed to be a direct precursor of **5**-$ACP_{II}$, which was disconfirmed by the results shown below.

### *4.3.4 Halogenation and Cyclopropane Ring Formation*

In assessing the role of Cur Hal in cyclopropane ring formation we considered the more evident role for the Hal in the analogous Jam pathway. In this system, chlorination evidently occurs on the pendant carbon generated by β-branching. Thus for Cur Hal, we reasoned that cyclopropane ring formation likely involves transient halogenation as in coronatine biosynthesis, where the chloride serves as a leaving group for intramolecular nucleophilic substitution (*19*). However, timing of the chlorination event remained to be established, and the identity of the Cur pathway ring-forming enzyme was not evident from bioinformatic analysis.

An important clue about the timing of chlorination at the β-branching carbon came from previous precursor-incorporation studies in curacin A biosynthesis. NMR data on curacin A labeling by [$^2H_3$,2-$^{13}$C]acetate indicated that the β-branching carbon that forms

cyclopropane was labeled by only one deuterium atom (α-isotope chemical shift at C20, 0.295) (*16*), which was previously interpreted as an anomalous result. Reconsideration of these data revealed that chlorination should occur before $ECH_2$ catalyzed decarboxylation. Otherwise, the pendant carbon atom would be labeled by either one or two deuterium atoms by 2:1 ratio (Figure 4-S6).

To identify the function of Cur Hal, we constructed the single-domain Hal and the tetradomain Hal-$ACP_3$ as N-terminal His-tagged proteins (Figure 4-S3). Cur Hal eluted as a dimer from an analytical size-exclusion column. Following His-tag removal by thrombin cleavage, the metal content of Hal was analyzed by inductively coupled plasma (ICP)-MS. After reconstitution with a mixture of metal ions and α-KG, more than 90% of Hal was bound exclusively to $Fe^{2+}$, which indicated that it functions as an $Fe^{II}$-dependent enzyme. Thus, anaerobic purification coupled with α-KG/$Fe^{II}$-reconstitution was employed, as was previously reported to retain optimal activities of α-KG-, $O_2$- and $Fe^{II}$-dependent halogenases (*19-21*).

We tested seven acyl-ACP substrates to establish the substrate identity for Cur Hal, including malonyl-$ACP_{IV}$, acetyl-$ACP_{IV}$, **1**-$ACP_{II}$, **2**-$ACP_{II}$, **3**-$ACP_{II}$, **4**-$ACP_{II}$ and **6**-$ACP_{II}$. Consistent with the [$^2H_3$,2-$^{13}$C]acetate precursor incorporation experiment noted above, we observed the formation of the mono-chlorinated species exclusively on **1**-$ACP_{II}$ to generate γ-Cl-**1**-$ACP_{II}$. The chlorinated product was confirmed by FTICR-MS and IRMPD (Figure 4-2a and 2c) analysis, and corroborated by GC-MS (Figure 4-3c, see below). Moreover, Cur Hal showed evident chirality preference toward the (*S*)-HMG-$ACP_{II}$ (**1**-$ACP_{II}$) substrate, as expected based on previous studies of Cur $ECH_1$ substrate selectivity (Figure 4-S5). In the absence of α-KG or $O_2$, no chlorinated product was detected with Cur Hal in the presence of HMG substrate (Figure 4-S5).

Next, we sought to investigate how chlorination of **1**-$ACP_{II}$ affects efficiency of the downstream reaction sequence with the HMG cassette enzymes. **1**-$ACP_{II}$ was converted to γ-Cl-**1**-$ACP_{II}$ by Hal and reacted sequentially with Cur $ECH_1$, $ECH_2$ and ER. $ECH_1$ dehydrated γ-Cl-**1**-$ACP_{II}$ and the γ-Cl-**2**-$ACP_{II}$ product was decarboxylated by $ECH_2$ to generate γ-Cl-**3**-$ACP_{II}$ (Figure 4-2a). Unexpectedly, no saturation product was observed

for Cur ER. Instead, we saw a 34-Dalton mass reduction from γ-Cl-**3**-$ACP_{II}$ (Figure 4-2a and 2d), demonstrating the elimination of chlorine in the product. This result suggests that the product could be **3**-$ACP_{II}$, **5**-$ACP_{II}$ or **6**-$ACP_{II}$. These experiments also revealed that the $ECH_1$/$ECH_2$-coupled dehydration and decarboxylation rate was significantly increased with the chlorinated substrate (Figure 4-S7), which might be due either to the electron-withdrawing effect of the γ-chlorine atom which stabilizes the negative charge on the intermediate of $ECH_2$ decarboxylation or to a more effective binding position of the chlorinated versus non-chlorinated substrate in the $ECH_2$ active site.

The experimental design to determine the final product in the presence of both Hal and HMG-cassette enzymes was streamlined by developing a one-pot reaction using Cur Hal-$ACP_3$, $ECH_1$, $ECH_2$ and ER. Cur (apo) Hal-$ACP_3$ was loaded with **1**-CoA, desalted, and mixed with Cur $ECH_1$, $ECH_2$ and ER in an anaerobic environment. The reaction was initiated by exposing the mixture to air. To confirm the structure of the final product, we cleaved the acyl group from the Hal-$ACP_3$ PPant arms with butylamine and compared the derivatives with the authentic standards by GC-MS. Direct correlation was made to a single species identified as the *cis*-2-methylcyclopropane-1-carboxyl compound (Figure 4-2e), demonstrating the formation of **5**-ACP by an unprecedented ER-catalyzed cyclopropanation reaction.

### *4.3.5 Regiochemical Control by $ECH_2$s*

Ascertaining the role of chlorination in the cyclopropane ring formation during curacin biosynthesis strongly suggested that a similar chlorination event occurs in the Jam pathway. Given the extraordinarily high similarity between Cur and Jam Hals (92% sequence identity), we surmised that the two pathways diverge after chlorination, resulting in differential catalytic processes in the terminal step to preserve a vinyl chloride moiety in jamaicamide. Compared to Cur and Jam Hals and $ECH_1$s, Cur and Jam $ECH_2$s and ERs have lower sequence identities (59% for $ECH_2$s and 65% for ERs), and likely function as key branch-point determinants. Accordingly, excised forms of Jam $ECH_2$ and ER were constructed, expressed and purified in the same way as the Cur enzymes (Figure 4-S3). We also expressed and purified Jam $ECH_1$ to generate substrate

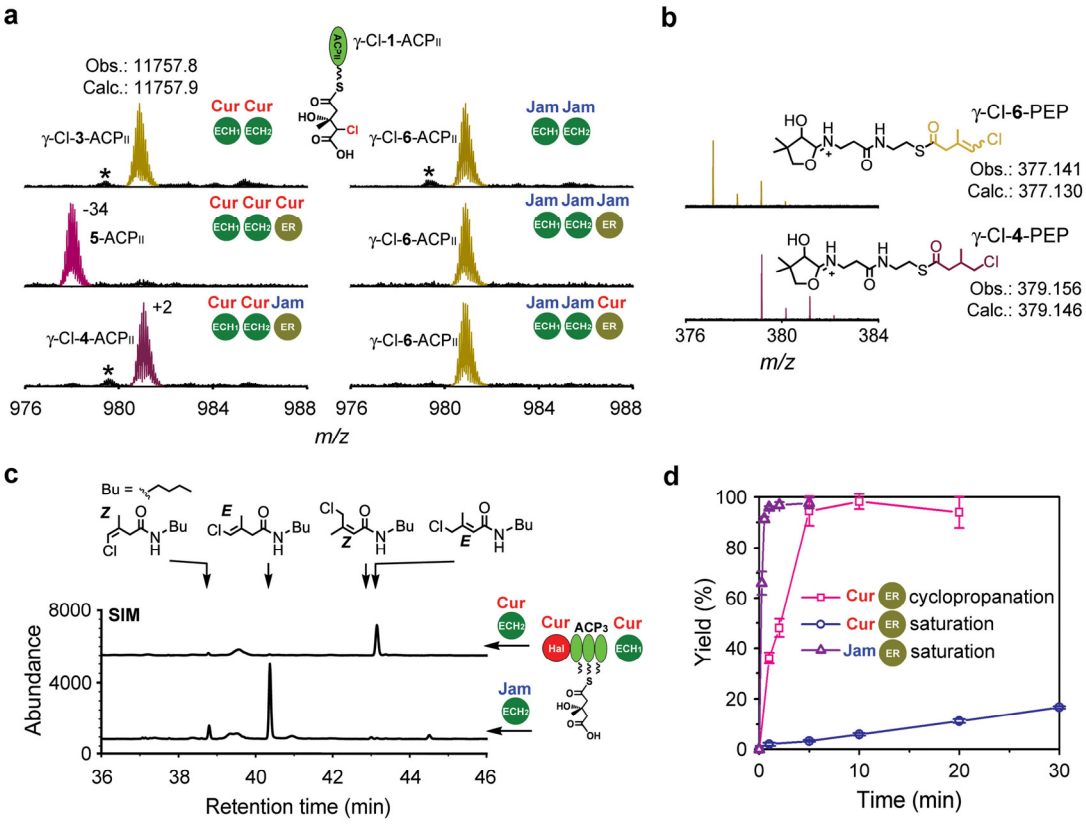



**Figure 4-3. Comparison of $ECH_2$s and ERs in Cur and Jam pathways**. **a**, Partial FTICR mass spectra for Cur and Jam $ECH_1$, $ECH_2$ and ER reactions with γ-Cl-**1**-$ACP_{II}$ substrate. The reactions were incubated at 30°C for 30 min. **b**, IRMPD spectra for the PEPs of γ-Cl-**6**-$ACP_{II}$ and γ-Cl-**4**-$ACP_{II}$. **c**, GC-MS analysis to identify the structures of Cur and Jam $ECH_2$ products. The chromatograms were recorded at selective ion mode (SIM) by monitoring 57, 117, 154 and 189 amu. The retention times of products were confirmed by co-injection with authentic standards. **d**, Comparison of catalytic efficiencies for cyclopropanation and saturation by Cur and Jam ERs. The product yields were measured by IRMPD-based quantification. **3**-$ACP_{II}$ was used as substrate for Cur ER saturation, and γ-Cl-**3**-$ACP_{II}$ was used as substrate for Cur ER cyclopropanation and Jam ER saturation. Assays were performed in triplicate.

for Jam $ECH_2$/ER reactions. We investigated the functions of Jam enzymes with Cur substrates starting from **1**-ACP or γ-Cl-**1**-$ACP_{II}$ to establish the key branch-point that controls the introduction of the β,γ vinyl chloride group.

For both **1**-$ACP_{II}$ and γ-Cl-**1**-$ACP_{II}$ substrates, Jam $ECH_1$ and $ECH_2$ catalyzed successive dehydration and decarboxylation steps as expected (Figure 4-3a, and Figure

4-S8a). However, when Jam ER was added, only ~20% of the saturated product was detected for the non-chlorinated substrate (derived from **1**-$ACP_{II}$, Figure 4-S8b). Remarkably, no mass change was observed for the corresponding chlorinated substrate (derived from γ-Cl-**1**-$ACP_{II}$), suggesting that the Jam $ECH_2$ product is not a substrate for Jam ER. Based on FTICR-MS and IRMPD results (Figure 4-3a, 3b, and Figure 4-S8), the product of Jam $ECH_2$ was consistent with either γ-Cl-**6**-$ACP_{II}$ (β,γ C=C; enoyl thioester) with a vinyl chloride group, γ-Cl-**3**-$ACP_{II}$ (α,β C=C; enoyl thioester) or a mixture of both.

To probe the key branching points in these related systems further, we selectively exchanged the Cur and Jam $ECH_1$/$ECH_2$/ER enzymes in coupled reactions with the γ-Cl-**1**-$ACP_{II}$ substrate. When Cur and Jam $ECH_1$s were switched in the coupled reaction, no change in the product profile was observed. However, when Jam $ECH_1$ and $ECH_2$ were coupled to Cur ER, only the Jam $ECH_2$ decarboxylation product was detected, in accord with the Jam $ECH_1$/$ECH_2$/ER reaction (Figure 4-3a). UV spectral comparison of Cur and Jam $ECH_2$ decarboxylation products revealed that their UV absorption patterns are slightly different between 250 and 280 nm (Figure 4-S9), which reflects isomeric α,β or β,γ C=C (enoyl thioester) functionality in the products (γ-Cl-**3**-$ACP_{II}$ and γ-Cl-**6**-$ACP_{II}$).

To determine the structures of the decarboxylation products, one-pot reactions using Cur Hal-$ACP_3$, $ECH_1$ and Cur or Jam $ECH_2$s were performed as described above. After butylamine cleavage of the products from the $ACP_3$, each butylamide derivative was analyzed by GC-MS and compared with the authentic standards. For the reaction including Cur $ECH_2$, the main product contained primarily an α,β C=C in the *E* configuration, with trace amounts of the β,γ C=C products (Figure 4-3c) quantified to be ~3% (Figure 4-4b, see below). In contrast, reactions using Jam $ECH_2$ showed a high regiochemical control to generate exclusively the β,γ C=C product, with ~85% in the *E* configuration and ~15% in the *Z* configuration (Figure 4-3c). The exclusive *E* configuration of the vinyl chloride C=C in jamaicamide natural products (*14*) suggests that the minor *Z* configuration product is likely due to utilization of the curacin substrate, which is less sterically hindered than the jamaicamide substrate. Notably, Jam $ECH_2$

decarboxylation had lower regiochemical control using the non-chlorinated substrate, and generated ~80% β,γ C=C and ~20% α,β C=C products, which further explains the partial enoyl reduced product observed following Jam $ECH_1$, $ECH_2$ and ER reactions with this substrate (Figure 4-S8b). Given the normal function of ER to catalyze only α,β C=C (enoyl thioester) saturation, the selective formation of β,γ C=C product by Jam $ECH_2$ renders Jam ER redundant in the biosynthesis of jamaicamides. In general, α,β C=C $ECH_2$ products are energetically preferred and frequently identified or predicted in other pathways (*13, 26, 30, 40-42*), except the pathways of pederin and its structural analogs (*33, 43*) that are predicted to generate β,γ C=C products (Figure 4-S10).

### *4.3.6* Functional Differentiation of ERs

Although Jam ER cannot function with the β,γ C=C thioester substrate, we sought to test whether it can catalyze cyclopropanation when presented with γ-Cl-**3**-$ACP_{II}$ bearing an α,β C=C. Thus, Cur $ECH_1$ and $ECH_2$ were coupled to Jam ER, but only the saturated product, γ-Cl-**4**-$ACP_{II}$, was observed (Figure 4-3a, 3b). Thus Jam ER has the activity of a canonical PKS ER, α,β C=C saturation.

The distinct functions of Cur and Jam ERs, despite the redundant nature of Jam ER within the Jam pathway, motivated us to compare the catalytic efficiencies of Cur ER cyclopropanation *vs*. Jam ER saturation of the chlorinated substrate, and the efficiencies of cyclopropanation of the chlorinated substrate *vs*. saturation of the non-chlorinated substrate by Cur ER. This was accomplished using time-course studies of catalytic efficiencies by measuring product yields under uniform reaction conditions. It was not possible to measure enzyme kinetic parameters ($k_{cat}$ and $K_M$) due to the tendency of ER to aggregate and the solubility limits of ACP-tethered substrates. Thus, γ-Cl-**3**-$ACP_{II}$ was employed to assess cyclopropanation by Cur ER, compared to reduction by Jam ER, and **3**-$ACP_{II}$ was used to compare reductive efficiency of the Cur and Jam ERs. IRMPD-based MS analysis (e.g. peak abundance of PPant ejection products (PEPs) (*38*)) provided a convenient method to quantify the yields of ER saturation products that correspond to a 2-Dalton mass change (Figure 4-S11). Similarly, Cur ER-catalyzed cyclopropanation was quantified by using **4**-$ACP_{II}$, prepared by loading (apo) $ACP_{II}$ with

**4**-CoA, as an internal standard for **5**-$ACP_{II}$.

We found that Jam ER saturation and Cur ER cyclopropanation of γ-Cl-**3**-$ACP_{II}$ are faster by ~400-fold and ~50-fold, respectively, than is Cur ER saturation of **3**-$ACP_{II}$ under identical experimental conditions (Figure 4-3d). For the **3**-$ACP_{II}$ substrate, Jam ER saturation is ~240-fold faster than is Cur ER saturation (Figure 4-S12). This comparison confirmed that Jam ER has retained a canonical α,β enoyl reduction function, in contrast to the unprecedented activity of Cur ER cyclopropanation.

### *4.3.7 Loss of Regiochemical Control by Mutation*

To understand the regiochemical control of $ECH_2$-catalyzed decarboxylation, the crystal structure of Cur $ECH_2$ was solved recently (*34*). Only three polar residues, $Lys^{86}$, $His^{240}$, and $Tyr^{82}$, are present in the nonpolar active site chamber, which is nevertheless water accessible (Figure 4-4a). As suggested by modeling with the non-chlorinated substrate and further supported by site-directed mutagenesis experiments, $Lys^{86}$ and $His^{240}$ are critical to catalysis (*34*). $His^{240}$ is conserved for all known $ECH_2$s in HMG enzyme cassettes, but $Lys^{86}$ and $Tyr^{82}$ are located in a hypervariable region (α2–loop–α3, residues 82-98) and are not conserved in Jam $ECH_2$ (*34*). As a follow-up analysis, Cur $ECH_2$ was modeled with the chlorinated substrate (Figure 4-4a), and the catalytic efficiencies of the wild type (WT) and mutants of Cur $ECH_2$ were compared using the chlorinated substrate γ-Cl-**2**-$ACP_{II}$. Moreover, we measured the ratios of the two possible decarboxylation products, γ-Cl-**3**-ACP and γ-Cl-**6**-ACP.

Similar to our previous assays (*34*), $ECH_1$/$ECH_2$-coupled dehydration and decarboxylation of γ-Cl-**1**-$ACP_{II}$ was analyzed to compare the catalytic efficiencies of WT and mutant Cur $ECH_2$s. It was immediately evident that their catalytic activities were significantly increased with the chlorinated substrate (Figure 4-4b) (*34*), possibly due to γ-Cl stabilization of the carbanion intermediate. As we observed previously for the non-chlorinated substrate under similar reaction conditions, the K86Q, K86A, H240Q and H240A variants had significantly decreased activities (Figure 4-4b). This suggests that $His^{240}$ and/or $Lys^{86}$ may draw the γ-carboxyl group away from the substrate in the transition state of $ECH_2$ decarboxylation through hydrogen-bond formation (Figure 4-4a).

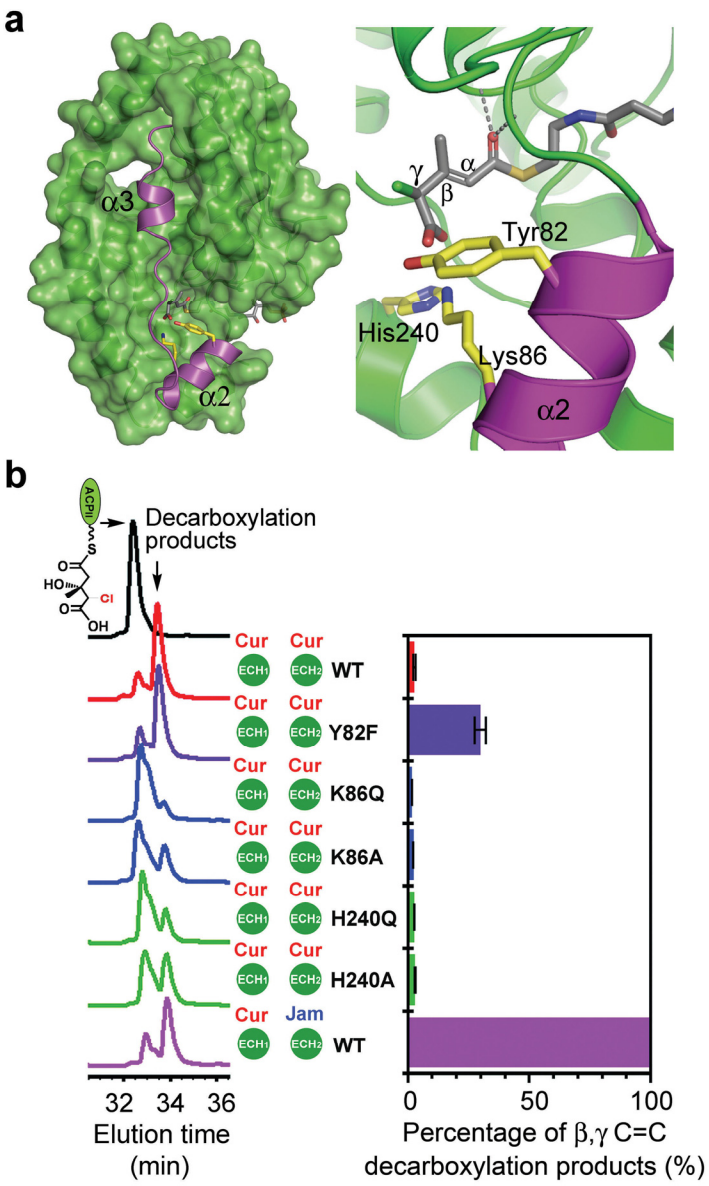



**Figure 4-4. Loss of Cur $ECH_2$-mediated regiochemical control by site-directed mutagenesis**. **a**, The hypervariable region (magenta) of Cur $ECH_2$ and the active site chamber modeled with the chlorinated substrate. The *S*-configuration of the γ-carbon is preferred based on modeling results. **b**, Activity and regiochemical control of $ECH_2$ WT and Cur $ECH_2$ mutants. γ-Cl-**1**-$ACP_{II}$ was used as the substrate for all reactions. (Left) HPLC analysis for $ECH_1$/$ECH_2$ coupled dehydration and decarboxylation. All reactions were quenched after 10 min incubation at 30°C. (Right) IRMPD-based quantification to measure the percent of β,γ C=C products. The coupled $ECH_1$/$ECH_2$ reactions were incubated for 45 min before treated with Jam ER for 45 min at 30°C.

The alanine substitutions at $Lys^{86}$ and $His^{240}$ had higher activity than the corresponding

glutamine substitutions (Figure 4-4b, left panel), possibly due to the small side chain of alanine allowing space for water molecules to stabilize the γ-carboxyl group or to donate a proton to the presumed enolate intermediate. Cur $ECH_2$ Y82F and Jam $ECH_2$ WT had similar activities to Cur $ECH_2$ WT, indicating that $Tyr^{82}$ is not essential to the decarboxylation step.

Next, we measured the ratio of α,β and β,γ C=C decarboxylation products to investigate whether the site-directed mutations can elucidate a basis for double bond regiochemical control by Cur $ECH_2$. Changes in the ratio of α,β and β,γ C=C products were assessed by measuring UV absorbance ratios ($A_{280nm}/A_{250nm}$, Figure 4-S9) for HPLC peaks corresponding to $ECH_2$ decarboxylation products (Figure 4-4b). Measured peak ratios for Cur $ECH_2$ WT, K86Q, K86A, H240Q and H240A are ~1.75, for Jam $ECH_2$ WT the ratio is 2.23, but for Cur $ECH_2$ Y82F it is 1.85. The intermediate value for Cur $ECH_2$ Y82F suggests a mixture of α,β and β,γ C=C products. These products can be distinguished directly by using Jam ER as a reagent to selectively reduce α,β C=C (Figure 4-3b, and Figure 4-S11) followed by IRMPD to quantitate the product ratios. The level of β,γ C=C product (γ-Cl-**6**-$ACP_{II}$) for Cur $ECH_2$ WT, K86Q, K86A, H240Q and H240A was ~3% of the total product formed, but was ~30% of total product formed by Cur $ECH_2$ Y82F (Figure 4-4b). Thus, replacing the $Tyr^{82}$ hydroxyl group with a hydrogen atom resulted in substantially reduced regioselectivity during Cur $ECH_2$ decarboxylation. As previously proposed, product regiochemistry is controlled by a protonation step leading to collapse of the presumed enolate intermediate (*34*). While the proton donor has not been identified, we propose that a hydrogen bond from the $Tyr^{82}$ hydroxyl group to the $Glu^{92}$ backbone carbonyl stabilizes the hypervariable loop, thereby blocking the entrance of water to protonate the α-C of the substrate to generate the β,γ C=C product (γ-Cl-**6**-$ACP_{II}$). Regiochemical control of the protonation step for Jam $ECH_2$ remains unclear since its crystal structure is not yet available. However, site-directed mutagenesis experiments described above suggest that this hypervariable region in $ECH_2$s plays an important role for regiochemical control.

## *4.4 DISCUSSION*

The Cur and Jam pathways enable us to witness the remarkable process of metabolic pathway evolution based on comparative biochemical analysis of their β-branching enzyme assemblies. What determines the evolution of chemical diversity in natural product pathways? This issue is particularly difficult to address in marine environments where host-microbe or predator-microbe interactions are poorly understood. In the case of a free living marine microorganism such as *Lyngbya majuscula*, it is clear that their large genomes contain high potential for natural product biosynthesis based on the presence of many distinct secondary metabolic pathways (*22, 44*). How these systems have developed within an individual cell, and how small or large fragments of DNA might participate in horizontal gene transfer remains unknown. However, that these events occur is evident by patterns of gene duplication, insertion and mutation, and moreover, by the close genetic and biochemical relationship between selected sets of biosynthetic components. Despite the structural relationships, subtle changes in amino acid sequences of only two members (e.g. $ECH_2$ and ER) of the 10 component β-branching enzyme system are ultimately responsible for the divergent chemical outcomes (Figure 4-1a). Thus, based on the genomic information and biochemical insights gained in the current study, a scenario of pathway evolution can be surmised. As a primary event, we propose that horizontal gene transfer resulted in insertion of an HMG gene assembly (Figure 4-1b) whose translation results in a full set of β-carbonyl reaction chemistry. This preformed metabolic system includes the Hal, discrete ACP (e.g. $ACP_{IV}$), KS, HCS, $ECH_1$, $ECH_2$ and ER, and specifies formation of a saturated β-methyl substituent during polyketide chain elongation (Figure 4-5a). The canonical example of this process can be found in the myxoveriscin A pathway where the full complement of enzymatic activities (except Hal) results in a pendant β-ethyl group at C16 in the 28-membered ring system (*30*).

Based on the initial biochemical characterization of the HMG-cassette multi-step process, variation in β-branching biochemistry is observed in a growing number of microbial secondary metabolic pathways. The functional group outcome is largely predictable from bioinformatic analysis. For example, mupirocin and bacillaene contain vinyl methyl substituents that reflect an HCS assembly devoid of an ER domain (*13, 41*),

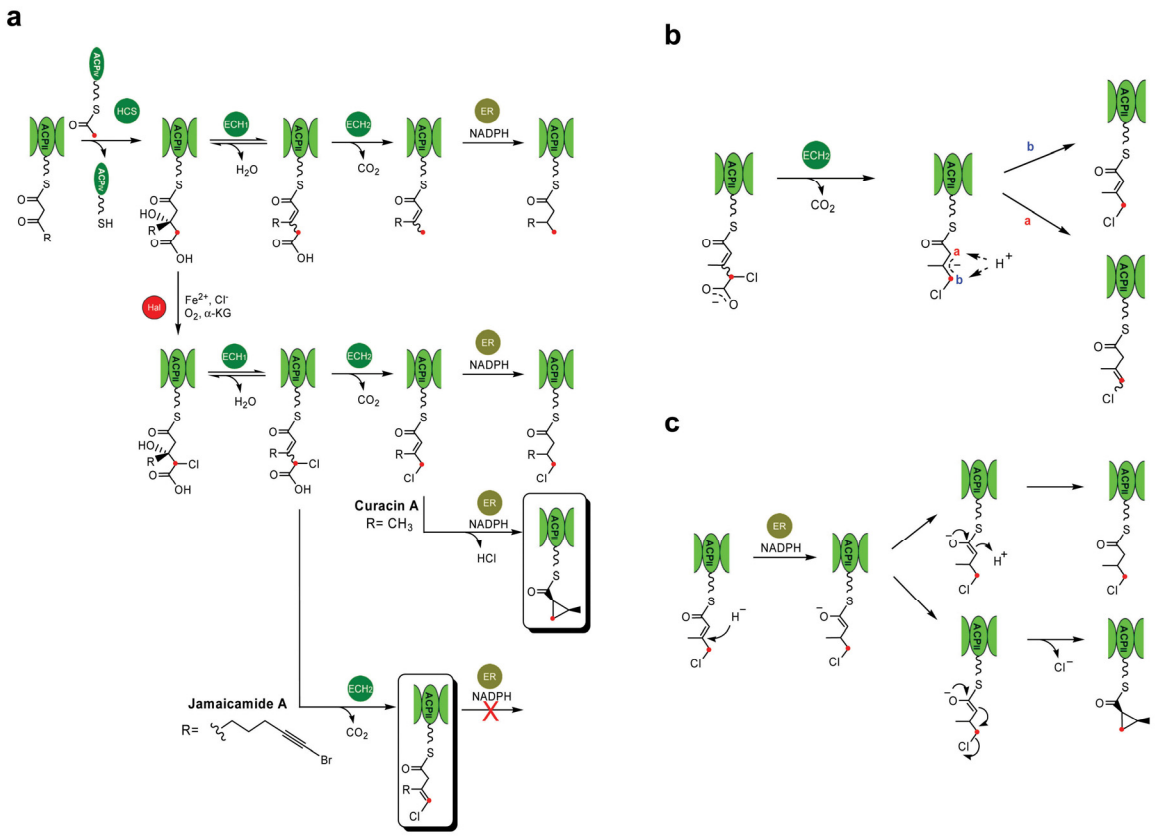



**Figure 4-5. Impact of enzyme assembly evolution on β-branching chemical diversity**. **a**, Proposed ancestral forms of the enzyme assemblies in Cur and Jam pathways. **b**, Differential regiochemical control by $ECH_2$. **c**, The functional diversification of ERs.

consistent with the architecture of their respective biosynthetic gene clusters. More unusual combinations of HCS assemblies are found in the leinamycin and bryostatin biosynthetic gene clusters (*31, 32*) that diverge considerably from the canonical example, where the formation of a β-branched HMG-ACP is followed by further elaboration of this intact subunit in the absence of dehydration/decarboxylation, or decarboxylation steps, respectively.

Where do Cur and Jam HMG assemblies reside in the spectrum of current examples leading to unique functional group chemistry? These two systems reveal clear evidence for two types of metabolic pathway evolution that result in unique biochemical functions and divergent paths for functional group assembly. Both HMG cassettes contain Hal domains that were evidently recruited, and embedded in a modular PKS to impart new

chemical diversity. Recent studies on this class of α-KG dependent non-haem halogenases have been reported as discrete enzymes within natural product pathways (*6, 19-22*), but this integrated domain represents an unprecedented example of pathway diversification. Although our initial analysis of the $ECH_1$ and $ECH_2$ catalyzed steps utilized HMG-ACP as substrate (*12*), we have demonstrated that the preferred substrate is γ-Cl-**1**-$ACP_3$. This halogenation step represents the point of divergence in the Cur and Jam pathways (Figure 4-5a) where substantial amino acid sequence variation is found in comparison of the respective $ECH_2$ and ER enzymes (Figure 4-1b). Pathway diversification is reflected in select amino acid sequence changes that ultimately control alternate double bond regiochemistry in the curacin and jamaicamide products. Regiochemical control is the product of a protonation step accompanying enolate collapse after $ECH_2$-mediated decarboxylation (Figure 4-5b). In this case, the ramifications of pathway evolution inform downstream steps in unexpected ways. Specifically, the Cur ER domain with its unusual placement outside a canonical PKS module has revealed itself to be a cyclopropanase, where hydride addition to the α,β-enoylthioester (γ-Cl-**3**-$ACP_3$) is followed by nucleophilic displacement of the chlorine atom leading to this highly strained and unusual functional group in secondary metabolism (Figure 4-5c). In contrast, our results demonstrate that the Jam ER domain, although still functional as a traditional reductase on the curacin α,β enoylthioester substrate, has evolved into a redundant enzyme on the alternative β,γ-enoylthioester substrate thus retaining the vinyl chloride functional group in jamaicamide. These parallel yet distinct systems demonstrate the mutability of enzymes within complex metabolic pathways, and reveal their metamorphic properties for creating chemical diversity in biologically active natural products.

## *4.5 MATERIAL AND METHODS*

The chemical synthesis of the butylamide derivatives as GC-MS authentic standards is outlined in Figure 4-S13. Butylamide derivatives, **8**, **10**, **12** and **14** were synthesized from carboxylic acids **7**, **9**, **11** and **13**, respectively.

**General Protocol A. Chemical synthesis and spectroscopic data of *N*-butyl-3-methylbutanamide (8).** A solution of carboxylic acid **7** (200 mg, 1.96 mmol)

in 20 mL of $CH_2Cl_2$ was treated with butylamine (160 mg, 2.19 mmol), HOBt (340 mg, 2.5 mmol), EDC hydrochloride (480 mg, 2.5 mmol) and *i*-$Pr_2NEt$ (1.1 mL, 6.5 mmol). The reaction mixture was stirred at room temperature for 15 h, concentrated *in vacuo*, and partitioned between cold (0˚C) 1 N HCl (100 mL) and $Et_2O$ (100 mL). The aqueous phase was further extracted with $Et_2O$ (2 x 100 mL). The combined organic extracts were washed with brine (150 mL), dried ($MgSO_4$), concentrated *in vacuo*, and purified by chromatography on $SiO_2$ (1:1, hexane/EtOAc) to provide **8** (250 mg, 80%) as a white powder: $^1$H NMR (400 MHz, $CDCl_3$) δ 3.15 (dd, *J* = 12.9, 7.1 Hz, 2 H), 1.96-2.09 (m, 1 H), 1.93 (d, *J* = 7.6 Hz, 2 H), 1.35-1.41 (m, 2 H), 1.22-1.28 (m, 2 H), 0.85 (d, *J* = 6.5 Hz, 6 H), 0.83 (t, *J* = 7.3 Hz, 3 H); MS (ESI) calculated for $[M+H]^+$ 158.15, found 158.14.

**Spectroscopic data for *N*-butyl-3-methylbut-2-enamide (10).** $^1$H NMR (400 MHz, $CDCl_3$) δ 5.46-5.50 (m, 1 H), 3.18 (dd, *J* = 13.5, 6.3 Hz, 2 H), 2.06 (s, 3 H), 1.74 (s, 3 H), 1.37-1.44 (m, 2 H), 1.22-1.31 (m, 2 H), 0.84 (t, *J* = 6.5 Hz, 3 H); MS (ESI) calculated for $[M+H]^+$ 156.13, found 156.12.

**Spectroscopic data for *N*-butyl-3-methylbut-3-enamide (12).** $^1$H NMR (400 MHz, $CDCl_3$) δ 4.87-4.89 (m, 1 H), 4.79-4.81 (m, 1 H), 3.15 (dd, *J* = 13.6, 6.5 Hz, 2 H), 2.86 (s, 2 H), 1.70 (s, 3 H), 1.34-1.42 (m, 2 H), 1.21-1.29 (m, 2 H), 0.83 (t, *J* = 7.3 Hz, 3 H); MS (ESI) calculated for $[M+H]^+$ 156.13, found 156.12.

**Spectroscopic data for (1*R*,2*S*)-*N*-butyl-2-methylcyclopropane-1-carboxamide (14) and its racemic mixture of *cis*- and *trans*-isomers (*cis,trans*-14).** The enantiomerically pure acid **13** was a gift from Timothy M. Ramsey (Novartis Institutes for Biomedical Research, Inc.). The racemic mixture of *cis*- and *trans*-isomers of **14** was generated from 2-methylcyclopropanecarboxylic acid. Data for **14**: $^1$H NMR (400 MHz, $CDCl_3$) δ 3.15 (dd, *J* = 13.7, 6.6 Hz, 2 H), 1.36-1.44 (m, 2 H), 1.21-1.31 (m, 2 H), 1.21-1.31 (m, 1 H), 0.96-1.04 (m, 1 H), 0.99 (d, *J* = 6.1 Hz, 3 H), 0.84 (t, *J* = 7.3 Hz, 3 H), 0.78-0.85 (m, 1 H), 0.44 (ddd, *J* = 7.7, 6.1, 3.6 Hz, 1 H); MS (ESI) calculated for $[M+H]^+$ 156.13, found 156.13. (Performed by Amol Kulkarni)

**Spectroscopic data and chemical synthesis of (*E*)-*N*-butyl-4-chloro-3-methylbut-2-enamide (18).** (*E*)-4-Bromo-3-methylbut-2-enoic

acid (**16**): To a solution of 3,3-dimethylacrylic acid (200 mg, 2.0 mmol) in $CCl_4$ (10 mL) was added freshly recrystallized NBS (391 mg, 2.2 mmol, 1.1 equiv) followed by benzoyl peroxide (24.2 mg, 0.05 equiv). The reaction mixture was heated at reflux for 1 h, cooled to room temperature, and filtered to remove succinimide. An aliquot was removed for $^1$H NMR analysis, and the remainder of the solution was poured into a separatory funnel and washed with $H_2O$ (2 x 20 mL) and brine (1 x 20 mL). The organic layer was dried ($Na_2SO_4$) and the solvent was removed *in vacuo* to afford a mixture (371 mg) of isomeric allylic bromides as a pale yellow oil. A solution of the crude oil (371 mg, 2.07 mmol) in water (2 mL) containing 82.8 mg (2.07 mmol) of NaOH was stirred at room temperature for 1 h and was then extracted with $CH_2Cl_2$ (2 x 2 mL). The aqueous layer was acidified with 6 N HCl (1 mL) and extracted again with $CH_2Cl_2$ (2 x 5 mL). The combined organic extracts were washed with brine (5 mL), dried ($Na_2SO_4$), and concentrated *in vacuo* to afford the desired carboxylic acid **16** (88 mg, 24%) as a pale yellow oil. This compound was carried on to the next step without further purification.

(*E*)-4-Chloro-3-methylbut-2-enoic acid (**17**): A solution of bromo acid **16** (360 mg, 2.01 mmol, 1.0 equiv) in $CH_2Cl_2$ (5 mL) was cooled to 0˚C, treated with *t*-butyl ammonium chloride (1.12 g, 4.02 mmol, 2.0 equiv) and stirred at 0˚C for 24 h. The mixture was concentrated *in vacuo* to afford a viscous pale yellow oil. Purification by chromatography on $SiO_2$ (2% MeOH/$CH_2Cl_2$ containing 0.5% AcOH) provided the desired acid **17** (215 mg, 79%) as a colorless oil.

(*E*)-*N*-Butyl-4-chloro-3-methylbut-2-enamide (**18**): To a solution of carboxylic acid **17** (40.0 mg, 0.297 mmol) in $CH_2Cl_2$ (0.5 mL) was added oxalyl chloride (0.377 g, 2.97 mmol, 10.0 equiv). The reaction mixture was heated at a gentle reflux for 2 h, allowed to cool to room temperature, and concentrated *in vacuo* to afford a viscous pale yellow oil. After addition of THF (0.5 mL), the solution was cooled to 0˚C, treated with a mixture of triethylamine (45 mg, 0.45 mmol, 1.5 equiv), butylamine (24 mg, 0.37 mmol, 1.1 equiv) and DMAP (1.8 mg, 0.015 mmol, 0.05 equiv), stirred at 0˚C for 15 min, and warmed to RT. The reaction mixture was stirred at room temperature for an additional 45 min and poured into a separatory funnel containing 1 M HCl (2 mL) and ether (2 mL). The organic layer was washed with water (1 x 2 mL) and brine (1 x 2 mL), dried ($Na_2SO_4$),

concentrated and purified by chromatography on $SiO_2$ to afford the desired butylamide derivative **18** (35 mg, 62%) as a pale yellow oil: IR (neat) 3296, 1633, 1548, 1266, 1182 $cm^{-1}$; $^1H$ NMR (300 MHz, $CDCl_3$) δ 5.88-5.86 (m, 1 H), 5.73 (br s, 1 H), 4.00 (s, 2 H), 3.29 (q, 2 H, *J* = 8.0 Hz), 2.21 (bs, 3 H), 1.50 (app p, 2 H, *J* = 7.5 Hz), 1.35 (app s, 2 H, *J* = 7.4 Hz), 0.92 (t, 3 H, *J* = 8.0 Hz); $^{13}C$ NMR (75 MHz, $CDCl_3$) δ 165.9, 146.6, 121.4, 50.2, 39.1, 31.6, 20.1, 16.3, 13.7; MS (ESI) calculated for $[M+H]^+$ 189.09, found 189.10. (Performed by Amol Kulkarni)

**Spectroscopic data and chemical synthesis of (*Z*)-*N*-butyl-4-chloro-3-methylbut-2-enamide (21).**

(*Z*)-4-Chloro-3methylbut-2-enoic acid (**20**): A solution of an (*E*,*Z*)-mixture of allylic bromides **19** (0.67 g, 3.74 mmol) in $CH_2Cl_2$ was treated at 0˚C with tetrabutyl ammonium chloride (2.08 g, 7.49 mmol), stirred at 0˚C for 24 h, and concentrated to afford a pale yellow viscous oil. Purification by chromatography on $SiO_2$ (1:99, $MeOH/CH_2Cl_2$ containing 0.5% AcOH) provided the chloroacid (0.402 mg, 79%) as a 1.2:1 mixture of (*E*,*Z*)- isomers. Further purification by SFC ($SiO_2$ column, 7% $MeOH/CO_2$) led to pure (*Z*)-isomer **20** (retention time = 3.12-3.17 min) (77 mg) as a colorless oil.

(*Z*)-*N*-Butyl-4-chloro-3-methylbut-2-enamide (**21**): To a solution of the (*Z*)-acid **20** (40.0 mg, 0.297 mmol) in $CH_2Cl_2$ (0.5 mL) was added oxalyl chloride (0.377 g, 2.97 mmol, 10.0 equiv). The reaction mixture was heated at reflux for 2 h, cooled to room temperature, and concentrated *in vacuo*. A solution of the viscous pale yellow oily residue in THF (0.5 mL) was cooled to 0˚C, treated with a mixture of triethylamine (45 mg, 0.45 mmol, 1.5 equiv), butylamine (24 mg, 0.33 mmol, 1.1 equiv) and DMAP (1.8 mg, 0.015 mmol, 0.05 equiv), stirred at 0˚C for 15 min and then allowed to warm to RT. The solution was stirred at room temperature for 45 min, and poured into a separatory funnel containing 1 M HCl (2 mL) and ether (2 mL). The organic layer was washed with water (1 x 2 mL) and brine (1 x 2 mL), dried ($Na_2SO_4$) and concentrated. The residue was purified by chromatography on $SiO_2$ (1% $MeOH/CH_2Cl_2$) to afford butylamide **21** (33 mg, 59%) as a pale yellow oil: IR (neat) 3424, 1691, 1631 $cm^{-1}$; $^1H$ NMR (300 MHz, $CDCl_3$) δ 5.69 (br s, 2 H) 4.69 (br s, 2 H), 3.29 (br t, 2 H, *J* = 6.6 Hz), 1.97 (d, 3 H, *J* = 1.2 Hz), 1.56-1.46 (m, 2 H), 1.42-1.29 (m, 2 H), 0.93 (t, 3 H, *J* = 7.2 Hz); $^{13}C$ NMR (75 MHz,

$CDCl_3$) δ 165.4, 146.8, 122.7, 42.8, 39.1, 31.6, 22.4, 20.1, 13.7; MS (ESI) calculated for $[M+H]^+$ 189.09, found 189.10. (Performed by Amol Kulkarni)

**Spectroscopic data and chemical synthesis of (*E*)-*N*-butyl-4-chloro-3-methylbut-3-enamide (25).**

(*E*)-4-Chloro-3-methylbut-3-enoic acid (**24**): A solution of $Cp_2ZrCl_2$ (1.46 g, 5.0 mmol, 1.0 equiv) in $CH_2Cl_2$ (5 mL) in a 50 mL 3-necked flask was stirred at 0˚C for 15 min and then treated with a solution of $AlMe_3$ (1.08 g, 15.0 mmol, 3.0 equiv) in $CH_2Cl_2$ (5 mL). The resultant pale yellow reaction mixture was stirred at 0˚C for 15 min, treated over 15 min with a solution of 3-butyn-1-ol (350 mg, 5 mmol) in $CH_2Cl_2$ (2 mL), slowly warmed to room temperature, stirred for 14 h, and cooled to -23˚C for 15 min before addition of NCS (800 mg, 6.0 mmol, 1.2 equiv). The mixture was stirred at -23˚C for 30 min, slowly warmed to room temperature and stirred for 15 min before the reaction was quenched by the slow, dropwise addition of 1 M HCl (20 mL). The solution was extracted with $CH_2Cl_2$ (2 x 20 mL). The organic layer was washed with water (1 x 20 mL) and brine (1 x 20 mL), dried ($Na_2SO_4$) and concentrated *in vacuo* to afford the alcohol **23** as a pale yellow oil. Purification by chromatography on $SiO_2$ (10% EtOAc/hexanes, followed by 20% EtOAc/hexanes) gave **23** (362 mg, 60%) as a colorless oil. A solution of alcohol **23** (30.1 mg, 0.25 mmol, 1.0 equiv) in acetone (1 mL) was cooled to 0˚C in an ice-bath for 15 min, treated with Jones reagent (0.25 mL, 0.63 mmol) and stirred at 0˚C for 20 min. After the addition of 2-propanol (400 μL), the reaction mixture was poured into a separatory funnel containing ether (5 mL) and water (5 mL). The aqueous layer was re-extracted with ether (2 x 5 mL). The combined organic layers were washed with brine (1 x 20 mL), dried ($Na_2SO_4$), and concentrated *in vacuo*. The resulting brown oil was purified by chromatography on $SiO_2$ (1% MeOH/$CH_2Cl_2$ containing 0.5% AcOH) to give the desired acid **24** (22.8 mg, 68%) as a pale yellow oil.

(*E*)-*N*-Butyl-4-chloro-3-methylbut-3-enamide (**25**): To a solution of the chloro acid **24** (34 mg, 0.25 mmol) in $CH_2Cl_2$ (0.5 mL) was added EDC•HCl (58 mg, 0.30 mmol, 1.2 equiv), triethylamine (38.4 mg, 0.379 mmol, 1.5 equiv) and butylamine (22.2 mg, 0.303 mmol, 1.2 equiv). The reaction mixture was stirred at room temperature for 14 h, poured into a separatory funnel containing 1 M HCl (1 mL) and ether (2 mL), and the organic

layer was washed with water (1 x 2 mL) and brine (1 x 2 mL) dried ($Na_2SO_4$) and concentrated *in vacuo.* The yellow oily residue was purified by chromatography on $SiO_2$ (10% EtOAc/hexanes, followed by 20% EtOAc/hexanes) to afford the desired amide **25** (20 mg, 41%) as a colorless oil: IR (neat) 3429, 1636, 1556 $cm^{-1}$; $^1H$ NMR (300 MHz, $CDCl_3$) δ 6.03-6.00 (m, 1 H), 5.61 (br s, 1 H), 3.30-3.19 (m, 2 H), 2.98 (s, 2 H), 1.85 (br d, 3 H, $J$ = 1.2 Hz), 1.54-1.43 (m, 2 H), 1.39-1.29 (m, 2 H), 0.92 (t, 3 H, $J$ = 7.2 Hz); $^{13}C$ NMR (75 MHz, $CDCl_3$) δ 169.1, 133.5, 116.6, 44.8, 39.4, 31.6, 20.0, 16.7, 13.7; MS (ESI) calculated for $[M+H]^+$ 189.09, found 189.10. (Performed by Amol Kulkarni)

**Spectroscopic data and chemical synthesis of (*Z*)-*N*-butyl-4-chloro-3-methylbut-3-enamide (28).**

(*Z*)-4-Chloro-3-methylbut-3-enoic acid (**27**). A solution of $Cp_2ZrCl_2$ (0.292 g, 1.0 mmol, 1.0 equiv) in 1,2-DCE (1 mL) in a 50 mL 3-necked flask was stirred at 0˚C for 15 min, treated with a solution of $AlMe_3$ (0.216 g, 3.0 mmol, 3.0 equiv) in 1,2-DCE (2 mL) and stirred at 0˚C for 15 min. To the reaction mixture was added a solution of 3-butyn-1-ol (70 mg, 1.0 mmol,1 equiv) in 1,2-DCE (1 mL) over 15 min. The mixture was slowly warmed to room temperature, stirred for 14 h, slowly heated to reflux and maintained at a reflux temperature for 3 d. The solution was cooled to 0˚C and subsequently to -23˚C for 15 min before addition of NCS (127 mg, 1.20 mmol, 1.2 equiv). The reaction mixture was stirred at -23˚C for an additional 30 min, slowly warmed to room temperature, stirred for 15 min, and quenched by a slow, dropwise addition of 1 M HCl (10 mL). The solution was extracted with ether (5 x 20 mL), and the combined organic layers were washed with brine (1 x 50 mL), dried ($Na_2SO_4$) and concentrated *in vacuo* to afford the desired alcohol **26** as a pale yellow oil. Purification of the crude compound by chromatography on $SiO_2$ (10% EtOAc/hexanes, followed by 20% EtOAc/hexanes) provided alcohol **26** (61 mg, 51%) as a pale yellow oil. A solution **26** (61 mg, 0.51 mmol, 1.0 equiv) in acetone (2 mL) was cooled to 0˚C for 15 min, treated with Jones reagent (0.51 mL, 1.28 mmol) and stirred for 20 min at 0˚C. The mixture was diluted with 2-propanol (400 μL), poured into a separatory funnel containing ether (5 mL) and water (5 mL), and the aqueous layer was washed with ether (2 x 5 mL). The combined organic layers were washed with brine (1 x 20 mL), dried ($Na_2SO_4$) and concentrated *in vacuo.* The brown oily residue was purified

by chromatography on $SiO_2$ (1% MeOH/$CH_2Cl_2$ containing 0.5% AcOH) to give the desired acid **27** (44.3 mg, 65%) as a pale yellow oil.

(*Z*)-*N*-Butyl-4-chloro-3-methylbut-3-enamide (**28**). To a solution of acid **27** (11.0 mg, 0.08 mmol, 1.0 equiv) in THF (1 mL) were added sequentially at 0˚C diphenylphosphinic chloride (19.3 mg, 0.082 mmol, 1.0 equiv), *N*-methylmorpholine (34.7 mg, 0.343 mmol, 4.2 equiv) and butylamine (5.9 mg, 0.08 mmol, 1.0 equiv). The reaction mixture was stirred at 0˚C for 1 h, concentrated and purified by chromatography on $SiO_2$ (20% EtOAc/hexanes) to give amide **28** (5.0 mg 32%) as a colorless oil: IR (neat) 3428, 1649, 1555, 1440 $cm^{-1}$; $^1H$ NMR (300 MHz, $CDCl_3$) δ 6.00 (app d, 1 H, $J$ = 1.5 Hz), 5.61 (br s, 1 H), 3.28-3.21 (m, 2 H), 3.16 (s, 2 H), 1.86 (d, 3 H, $J$ = 1.2 Hz), 1.53-1.43 (m, 2 H), 1.40-1.25 (m, 2 H), 0.92 (t, 3 H, $J$ = 7.2 Hz); $^{13}C$ NMR (75 MHz, $CDCl_3$) δ 168.6, 134.0, 114.9, 40.4, 39.4, 31.6, 21.4, 20.0, 13.7; MS (ESI) calculated for $[M+H]^+$ 189.09, found 189.10. (Performed by Amol Kulkarni)

**Spectroscopic data and chemical synthesis of *N*-butyl-4-chloro-3-methylbutanamide (31).**

*N*-Butyl-4-chloro-3-methylbutanamide (**31**). To a solution of lactone **29** (100 mg, 1.0 mmol, 1.0 equiv) in thionyl chloride (0.13 g, 1.0 equiv) was added freshly fused $ZnCl_2$ (6.8 mg, 0.05 mmol, 0.05 equiv). The reaction mixture was slowly heated to reflux, maintained at reflux for 8 h, and then slowly cooled to room temperature. The mixture was concentrated *in vacuo* and a solution of the brown oily residue in THF (0.5 mL) was cooled to 0˚C, treated with a mixture of triethylamine (0.15 g, 0.15 mmol, 1.5 equiv) and butylamine (80 mg, 1.1 mmol, 1.1 equiv), stirred at 0˚C for 30 min, warmed to room temperature, stirred for 30 min, and poured into a separatory funnel containing 1 M HCl (2 mL) and $CH_2Cl_2$ (2 mL). The organic layer was washed with brine (1 x 2 mL), dried ($Na_2SO_4$) and concentrated *in vacuo* to afford a pale brown oil that was purified by chromatography on $SiO_2$ (20% EtOAc/hexanes followed by 50% EtOAc/hexanes) to give amide **31** (70.7 mg, 37%) as a yellow oil: IR (neat) 3422, 1644, 1559, 1459, 1379 $cm^{-1}$; $^1H$ NMR (300 MHz, $CDCl_3$) δ 5.92 (br s, 1 H), 3.59-3.48 (m, 2 H), 3.24 (app q, 2 H, $J$ = 6.6 Hz), 2.50-2.30 (m, 2 H), 2.10 (dd, 1 H, $J$ = 14.1, 7.2 Hz), 1.52-1.43 (m, 2 H), 1.37-1.27 (m, 2 H), 1.04 (d, 3 H, $J$ = 6.6 Hz), 0.91 (t, 3 H, $J$ = 7.2 Hz); $^{13}C$ NMR (75

MHz, $CDCl_3$) δ 171.3, 50.6, 40.6, 39.3, 32.6, 31.6, 20.0, 17.6, 13.7; MS (ESI) calculated for $[M+H]^+$ 191.11, found 191.10. (Performed by Amol Kulkarni)

**Chemical synthesis of 6-CoA.** The chemical synthesis of **6**-CoA was described in our previous work (*12*).

**Enzymatic synthesis of 1-CoA.** **1**-CoA was enzymatically generated by HMG-CoA reductase as previously described (*12, 45*). The reaction mixture was separated by a self-packed DEAE Sepharose column (1 cm x 30 cm, GE Healthcare)) equilibrated with 0.1 M $HCO_2NH_4$, pH 4.4. **1**-CoA was eluted with 0.4 M $HCO_2NH_4$, pH 4.4 (*45*). **1**-CoA was further purified using XBridge Preparative C18 column (10 x 250 mm, 5 μm) and an elution gradient of 3-60% $CH_3OH/H_2O$ (10 mM $NH_4OAc$). The fractions were pooled and lyophilized, and **1**-CoA purity was measured by HPLC to be >98%.

**Enzymatic synthesis of 2-CoA.** **2**-CoA was enzymatically generated from **1**-CoA by Cur $ECH_1$ as previously described (*34*).

**Bacterial strains, media and culture conditions.** *Escherichia coli* DH5α MCR (Invitrogen) was used for DNA propagation. *Escherichia coli* BL21 (DE3) transformed with the derivatives of pET20b, pET24b, pET28b, and pET41a (Novagen) were used for protein overexpression in Luria-Bertani (LB) medium. Ampicillin (100 μg/mL), carbenecillin (100 μg/mL), kanamycin (50 μg/mL), chloramphenicol (25 μg/mL), and apramycin (50 μg/mL) were used for the corresponding plasmid construct resistance marker selection in *E. coli* cultures.

**Plasmid construction and site mutagenesis.** The primers for the plasmid construction were listed in Table S1. The expression plasmids for CurB $ACP_{IV}$, CurE $ECH_1$ and CurF $ECH_2$ were constructed in our previous work (*12*). CurA Hal, Hal-$ACP_3$ and CurF ER genes were amplified from the cosmid pLM54 (*16*). JamI $ECH_1$, JamJ $ECH_2$ and JamJ ER genes were amplified from cosmid pJam3 (*14*). CurA Hal and Hal-$ACP_3$ genes were inserted into pET28b plasmid using *Nde*I/*Xho*I restriction sites. JamI $ECH_1$ gene was inserted into pET24b plasmid using *Nde*I/*Xho*I restriction sites. JamJ $ECH_2$ gene was inserted into pET20b plasmid using *Nde*I/*Xho*I restriction sites. CurF ER and JamJ ER

genes were inserted into pET41a plasmid using *Sac*II/*Xho*I restriction sites. The CurA $ACP_3$, $ACP_I$, $ACP_{II}$ and $ACP_{III}$ expression plasmids derived from a synthetic $ACP_3$ gene were provided by Professor Christopher Walsh and Dr. Christopher Calderone at Harvard Medical School. CurA $ACP_3$ (1946-2248) gene was inserted into pET28a plasmid using *Nde*I/*Xho*I restriction sites. CurA $ACP_I$ (1946-2031), $ACP_{II}$ (2057-2146) and $ACP_{III}$ (2161-2248) genes were inserted into pET29a plasmid using *Nde*I/*Xho*I restriction sites. CurF $ECH_2$ mutants, H240A, H240Q, K86A, K86Q and Y82F, were made in our previous work (*34*). All the constructs and mutations were verified by DNA sequencing.

**Protein overexpression.** The protein expression conditions for CurB $ACP_{IV}$, CurE $ECH_1$, CurF $ECH_2$ and the corresponding mutants were described in our previous work (*12, 34*). **A) CurA (apo) $ACP_3$, (apo) $ACP_I$, (apo) $ACP_{II}$, (apo) $ACP_{III}$, JamI $ECH_1$,** and **JamJ $ECH_2$:** *E. coli* BL21 (DE3) was transformed by the pET28a::$ACP_3$ (pCC111) plasmid to overexpress the N-terminal His-tagged protein, and by pET29a::$ACP_I$ (pCC112), pET29a::$ACP_{II}$ (pCC113), pET29a::$ACP_{III}$ (pCC114), pET24b::Jam $ECH_1$ and pET20b::Jam $ECH_2$ to overexpress the C-terminal His-tagged proteins. Cells were grown at 35˚C to an OD (590 nm) = 0.5-0.6, and then cooled to 18˚C prior to addition of 1 mM isopropyl-β-D-galactopyranoside (IPTG). The cultures were grown at 18˚C for another 12-15 h before harvesting. **B) CurA Hal** and **(apo) Hal-$ACP_3$:** *E. coli* BL21 (DE3) was transformed by pET28b::Cur Hal and pET28b::Cur Hal-$ACP_3$ plasmids to overexpress the N-terminal His-tagged proteins. Cells were grown at 30˚C to an OD (590 nm) = 0.5-0.6, and then cooled to 15˚C prior to the addition of 1 mM IPTG. The cultures were grown at 15˚C for another 18-20 h before harvesting. **C) CurF ER** and **JamJ ER:** *E. coli* BL21 (DE3) was transformed by pET41b::Cur ER and pET41b::Jam ER to overexpress the N-terminal GST-tagged and C-terminal His-tagged proteins. Cells were grown at 30˚C to an OD (590 nm) = 0.5-0.6, and then cooled to 18˚C prior to the addition of 1 mM IPTG. The cultures were grown at 18˚C for another 12-15 h before harvesting.

**Protein purification.** Protein purifications were performed at 4˚C. In general, the first step Ni-affinity purifications for all His-tagged proteins were performed under the same conditions. *E. coli* cells were harvested by centrifugation (5,000 g, 20 min, 4˚C), resuspended in ice cold lysis buffer A (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 10

mM imidazole, 20% glycerol) and disrupted by sonication on ice. The cell debris was removed by centrifugation at 15,000 g for 50 min. The supernatant was gently removed and loaded onto the 5 mL HisTrap column (GE Healthcare) preequilibrated with lysis buffer A. The resin was washed successively with ~10 column volumes of washing buffer B (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol) to remove nonspecifically bound contaminants. Bound proteins were eluted with imidazole by a linear gradient of the elution buffer C (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 250 mM imidazole, 20% glycerol). The fractions of the eluate were checked for purity by SDS-PAGE, pooled and concentrated using Amicon Ultra-15 (30 kDa, 10 kDa or 5 kDa) centrifugal devices (Millipore). The follow-up purification and buffer exchange for different proteins are described below:

**A) CurA (apo) $ACP_3$, (apo) $ACP_I$, (apo) $ACP_{II}$, (apo) $ACP_{III}$, CurB (apo) $ACP_{IV}$, CurF $ECH_2$ WT** and **the mutants, CurF ER, JamI $ECH_1$, JamJ $ECH_2$** and **JamJ ER:** the concentrated eluate was loaded onto HiPrep 26/10 Desalting column (GE Healthcare) equilibrated with the storage buffer D (50 mM PBS buffer, pH 7.5, 200 mM NaCl, 20% glycerol), The fractions were pooled, concentrated, flash-frozen in 50-200 μL aliquots by liquid $N_2$, and stored at -80˚C. **B) CurE $ECH_1$:** CurE $ECH_1$ is prone to precipitate after its elution from HisTrap column. So the eluted protein was immediately desalted by PD10 column (GE Healthcare) equilibrated with the storage buffer D. The protein was flash-frozen in 100 μL aliquots by liquid $N_2$ and stored at -80˚C.

**Anaerobic purification of Cur Hal and (apo) Hal-$ACP_3$.** The $Fe^{II}$-, α-KG- and $O_2$-dependent halogenases were reported to be sensitive to $O_2$, and anaerobic purification was performed to remain their activities (*19, 20*). The purification was performed under inert atmosphere by using ÄKTA FPLC (GE Healthcare) companied with a glove box (Coy Laboratory Products), which is similar to the system previously described (*46*). All the buffers were sparged with argon and equilibrated in glove box for two days. *E. coli* cells were resuspended in buffer A and disrupted by sonication on ice. Then the cell lysate was equilibrated with argon for ~1 hour, transferred to gas-tight tubes in glove box, and centrifuged at 15,000 g for 50 min. The Ni-affinity purification for Hal and Hal-$ACP_3$ was performed as described above. After elution from HisTrap column, the

proteins were exchanged into buffer D using HiPrep 26/10 Desalting column. The N-terminal His-tag of Hal was removed by thrombin for metal analysis (*19, 20*). The biotinylated thrombin (Novagen) was added into Hal in buffer D, and the reaction was incubated at room temperature for overnight to achieve a complete His-tag cleavage. The biotinylated thrombin was removed from the reaction mixture by using 1 mL HiTrap Streptavidin HP column (GE Healthcare). The flow-through was loaded onto HiLoad 26/60 Superdex 200 column (GE Healthcare) equilibrated with the storage buffer D. Cur Hal was eluted as a dimer from size-exclusion columns. The fractions were pooled, concentrated, flash-frozen in 50 μL aliquots by liquid $N_2$, and stored at -80˚C.

**Metal analysis of Cur Hal.** To determine which metal bound to Hal, His-tag cleaved Cur Hal was anaerobically reconstituted with 1 mM α-KG and a metal mixture containing 25 μM $Fe^{2+}$, $Co^{2+}$, $Ni^{2+}$, $Cu^{2+}$ and $Zn^{2+}$ for 10 min before desalted by PD10 column equilibrated with buffer E (50 mM Tris-HCl buffer, pH 7.5, 150 mM NaCl, 10% glycerol). The metal content of Cur Hal was measured by ICP-MS (Finnigan). The buffer E was applied as blank for analysis. For ~2.0 μM Cur Hal, the metal concentrations were measured to be 1.81 ± 0.09 μM Fe (blank 0.09 ± 0.02), < 0.0031 μM Co (blank < 0.0027), < 0.0017 μM Ni (blank < 0.0015), < 0.0019 μM Cu (blank < 0.0017), and 0.067 ± 0.003 μM Zn (blank 0.039 ± 0.003). Based on this result, the reconstitution of with α-KG and $Fe(NH_4)_2(SO_4)_2$ were performed for its activity assay described below.

**Substrate loading onto (apo) ACPs.** The ACP-linked substrates were generated by loading the corresponding CoA substrates onto (apo) ACPs by using recombinant *Bacillus subtilis* Sfp (*12, 47*). Typically, 500 μM acyl-CoA and 50 μM (apo) ACP were incubated with 2 μM Sfp, and 10 μM $MgCl_2$ in 50 mM Tris-HCl buffer, pH 8.1, at room temperature for ~2 h. Reaction mixtures were desalted by PD10 column equilibrated with the buffer E. The desalted (holo) ACPs were concentrated by using Amicon Ultra-4 (5 kDa, Millipore), flash-frozen in 10-50 μL aliquots in liquid $N_2$, and stored at -80˚C. For substrate loading onto (apo) Hal-$ACP_3$, the reagents were equilibrated with argon prior to use. The substrate loading onto (apo) Hal-$ACP_3$ was manipulated in glove box. The substrate-loaded ACP samples were analyzed by reverse-phase HPLC using a Jupiter C4 column (250 x 2.0 mm, 5 μm, 300 Å, Phenomenex), and a linear elution gradient from 5%

to 90% of $CH_3CN$ (0.1% $CF_3CO_2H$)/$H_2O$ (0.1% $CF_3CO_2H$).

**Cur Hal activity assays & enzymatic generation of γ-Cl-1-$ACP_{II}$.** Cur Hal acitivity was tested by using $ACP_{IV}$ or $ACP_{II}$-linked substrates including malonyl-$ACP_{IV}$, acetyl-$ACP_{IV}$, **1**-$ACP_{II}$, **2**-$ACP_{II}$, **3**-$ACP_{II}$, **4**-$ACP_{II}$ and **6**-$ACP_{II}$. The chlorination product was only detected for **1**-$ACP_{II}$. As such, the follow-up chlorination assays were performed by incubating the enzyme with **1**-$ACP_{II}$ or **1**-$ACP_3$. Typically, 200 μl reaction mixture containing 50 μM **1**-$ACP_{II}$ or **1**-$ACP_3$, 5 μM Cur Hal, 50 μM fresh $Fe(NH_4)_2(SO_4)_2$, and 0.5 mM α-KG in 50 mM Tris-HCl buffer (pH 7.5) was prepared in glove box. The reaction was initiated by exposing the reaction mixture to air, and incubated at 30˚C for 2 h to achieve a full conversion to γ-Cl-**1**-$ACP_{II}$. The γ-Cl-**1**-$ACP_{II}$ was served as substrate for $ECH_1$/$ECH_2$/ER reactions. FTICR-MS and IRMPD were employed to detect the products.

**α-KG and $O_2$ dependence of Hal chlorination.** The α-KG and $O_2$ dependence of Cur Hal chlorination was investigated by incubating the enzyme with **1**-$ACP_{II}$ in the absence of α-KG or $O_2$ under the uniform reaction condition described above. For Hal chlorination without $O_2$, the reaction mixture was prepared in glove box, and capped during its incubation.

**$ECH_1$ and $ECH_2$ acitivity assays**. $ECH_1$ and $ECH_2$ assays were performed as previously described (*34*). Briefly, ~50 μM **1**-$ACP_{II}$ or γ-Cl-**1**-$ACP_{II}$ was added with 1 μM $ECH_1$ or $ECH_1$/$ECH_2$ in 50 mM Tris-HCl buffer (pH 7.5), and incubated at 30˚C. After quenched by addition of 10% formic acid, the reactions were analyzed by reverse-phase HPLC using Jupiter C4 column. FTICR-MS and IRMPD were applied to detect the products.

**ER activity assays.** The chlorinated substrate, γ-Cl-**3**-$ACP_{II}$, was generated from γ-Cl-**1**-$ACP_{II}$ by Cur $ECH_1$/$ECH_2$, and the non-chlorinated substrate, **3**-$ACP_{II}$ was generated by loading (apo) $ACP_{II}$ with **3**-CoA. Typically, ER reactions were performed by incubating ~50 μM γ-Cl-**3**-$ACP_{II}$ or **3**-$ACP_{II}$ with 1 μM ER and 0.5 mM NADPH in 50 mM Tris-HCl buffer (pH 7.5) at 30˚C. Alternatively, ER reaction was coupled with $ECH_1$/$ECH_2$ dehydration and decarboxylation by serving γ-Cl-**1**-$ACP_{II}$ or **1**-$ACP_{II}$ as the

substrate. FTICR-MS and IRMPD were applied to detect the products.

**One-pot reactions with Cur Hal-$ACP_3$ or $ACP_3$-linked substrate.** The one-pot reactions were conducted to confirm the products of $ECH_1/ECH_2$, Hal/$ECH_1/ECH_2$, $ECH_1/ECH_2$/ER, and Hal/$ECH_1/ECH_2$/ER coupled reactions. The $ACP_3$-linked HMG substrates were generated by loading Cur (apo) Hal-$ACP_3$ and $ACP_3$ with **1**-CoA as described above. The substrate loading of (apo) Hal-$ACP_3$ was manipulated in glove box. Typically, the one-pot reactions were performed by incubating ~50 μM $ACP_3$ or Hal-$ACP_3$-linked HMG substrate with ~10 μM enzymes and their corresponding cofactors at 30˚C for 5 min. The reactions were initiated by exposing the reaction mixture to $O_2$, and quenched by addition of 10% formic acid. Hal-$ACP_3$ and $ACP_3$ were purified from their reaction mixtures by reverse-phase HPLC before subjected to butylamine aminolysis as described below.

**Butylamine aminolysis.** To determine the structures of products linked to the $ACP_3$ and Hal-$ACP_3$ PPant arms, the acyl groups were cleaved by butylamine aminolysis to generate the butylamide derivatives, which were analyzed by GC/MS (*39*) and compared with the authentic standards. To remove contaminants, the $ACP_3$ and Hal-$ACP_3$ were purified from their reaction mixtures by reverse-phase HPLC using the Source 15PRC column. The proteins were eluted using a linear gradient from 30% to 70% of $CH_3CN$ (0.1% $CF_3COOH$)/$H_2O$ (0.1% $CF_3COOH$), and pooled in glass vials before lyophilization. The aminolysis reaction was performed by adding 160 μl $H_2O$, 100 μl hexane and 40 μl butylamine into ~5 nmol lyophilized $ACP_3$ or Hal-$ACP_3$ samples. The reaction mixture was incubated at 30˚C for 30 min, quenched by 66 μl 12 M HCl, and extracted with 2 x 2 mL ethyl acetate. The extracts were dried under nitrogen, and the butylamides were redissolved in 100 μl hexane before GC/MS analysis.

**GC/EI-MS analysis**. The samples and authentic standards were analyzed by a Hewlett Packard 6890 gas chromatograph equipped with a 5973 mass selective quadrupole detector. The butylamides were separated on a DB-VRX (Agilent J&W) capillary column (60 m x 253 μm x 1.40 μm), which was operated with helium-carrier gas and splitless injection. Both the injector and detector temperatures were set as 250˚C. After initial 6

min at 45˚C, the oven temperature was raised to 140˚C at 8˚C/min and held for 10 minutes, and then to 225˚C at 30˚C/min and held for 25 minutes. Total ion chromatograms were recorded using a mass range of 35-270 amu, and the selective ion chromatograms were recoded by monitoring 2-3 most abundant masses plus the parent masses of target compounds.

**Analysis of ACP samples by electrospray ionization (ESI)-FTICR-MS**. The observed and calculated masses for all $ACP_{II}$ samples are listed in Table S2. The ESI-FTICR-MS samples were prepared by separating $ACP_{II}$ from the reaction mixtures using the Source 15PRC reverse phase column as previously described (*47*). All samples were freshly prepared and analyzed with an actively shielded 7 Tesla quadrupole- FTICR mass spectrometer (APEX-Q, Bruker Daltonics, Billerica, MA). Target analytes in electrospray solution (1:1 $CH_3CN$:$H_2O$ with 0.1% HCOOH) were directly infused into an electrospray ionization (ESI) source (Apollo II, Bruker Daltonics) operating in positive ion mode at a flow rate of 70 μL/h and a voltage of - 3.8 kV. A counterflow of hot (240°C) nitrogen gas was applied to assist desolvation of ESI droplets. Multiply protonated ions generated by ESI were externally accumulated in a hexapole and transferred via high voltage ion optics to the ICR cell for analysis. For IRMPD, precursor ions were mass-selectively accumulated in the hexapole with a 5-10 m/z quadrupole isolation window, transferred to the ICR cell, and irradiated for 100-300 ms by 10.6 μm photons at 10 W laser power (25 W $CO_2$ laser, Synrad, Mukilteo, WA). All data were acquired with XMASS software (version 6.1, Bruker Daltonics) in broadband mode from m/z = 200 to 2000 with 512k data points and summed over 10-30 scans. Mass spectra were analyzed with the MIDAS analysis software (*48*). For accurate mass determination, apomyoglobin (Sigma, St. Louis, MO) peaks on charge state of 19-20 (apomyoglobin was spiked into the ESI solution prior to analysis), or known *y*-type product ions ($y_{22}$) generated from IRMPD of the same protein were used as internal calibrants. Once the exact mass of a protein had been determined, its two most abundant charge states were selected as internal standards for further calibration (apomyoglobin was not spiked into all reactions). All frequency-to-m/z calibrations were performed with a two-term calibration equation (*49*). (Performed by Bo Wang and Liangcai Gu)

**IRMPD-based quantification.** Peak abundances of PPant ejection products (PEPs) generated by IRMPD (*38*) were measured to quantify the yields of the ER catalyzed α,β C=C saturation and cyclopropanation reactions for time-course studies, or site mutagenesis assays. In ER α,β C=C saturation reaction, an addition of two hydrogens to the substrate is unlikely to affect ESI efficiency. Furthermore, products and substrates were observed with the same charge state distribution patterns, allowing a single charge state to be used for analysis (*50*). For Cur ER cyclopropanation, quantification of **5**-$ACP_{II}$ was achieved by using **4**-$ACP_{II}$ as an internal standard. No desaturation product of **4**-$ACP_{II}$, which would interfere with **5**-$ACP_{II}$ measurement, was detected. The ratio of these acyl-$ACP_{II}$ species with 2 Dalton mass difference can be straightforwardly measured by analyzing their PEPs. With IRMPD, quantification results from a single charge state (+12) were identical to those based on an average of the entire charge state distribution. Thus, we chose to utilize the +12 charge state for IRMPD quantification analysis. The ratio of PEPs with 2 Dalton mass difference was calculated by measuring the abundance of the n and n+2 product ion peaks (the n+2 peak abundance was adjusted by subtracting the natural n+2 isotopic abundance for peak n). For saturation reactions, yields were calculated by measuring the ratios of substrates with respect to products whereas, for cyclopropanation reactions, yields were calculated by measuring the ratios of products with respect to the internal standard. (Performed by Bo Wang and Liangcai Gu)

**Time-course studies of ER saturation and cyclopropanation.** The time-course studies were performed to compare the catalytic efficiencies of Cur ER cyclopropanation and saturation, and Jam ER saturation. γ-Cl-**3**-$ACP_{II}$ was applied as the substrate for Cur ER cyclopropanation and Jam ER saturation, and **3**-$ACP_{II}$ as the substrate for Cur and Jam ER saturation. All the reactions were performed in triplicate under the uniform reaction condition. ~200 μM γ-Cl-**3**-$ACP_{II}$ or **3**-$ACP_{II}$ was incubated with 2 μM Jam or Cur ER, and 1 mM NADPH in 50 mM Tris-HCl buffer (pH 7.5) at 30˚C. For Cur ER cyclopropanation, ~100 μM **4**-$ACP_{II}$, which was generated from (apo) $ACP_{II}$ and **4**-CoA, was added as the internal standard for **5**-$ACP_{II}$. No desaturation product of **4**-CoA was observed under the tested conditions. After certain amounts of reaction time, ~20 μl

aliquots were removed from the reaction mixture, added with 50 µl 10% formic acid to quench the reaction, and flash-frozen by liquid $N_2$. The $ACP_{II}$ samples were purified by reverse-phase HPLC, and immediately subjected to the IRMPD-based quantification analysis.

**Quantification of regiochemical products of the $ECH_2$ WT and mutants.** The ratios of α,β and β,γ C=C products of the $ECH_2$ decarboxylation were measured for Cur $ECH_2$ WT and mutants, and Jam $ECH_2$ WT. The products were generated from γ-Cl-**1**-$ACP_{II}$ by $ECH_1$/$ECH_2$ coupled dehydration and decarboxylation. 50 µM γ-Cl-**1**-$ACP_{II}$ was incubated with 2 µM Cur $ECH_1$ and 2 µM Cur or Jam $ECH_2$ in 50 mM Tris-HCl buffer (pH 7.5) at 30˚C for 45 min. Next, the reaction mixtures were treated with 2 µM Jam ER with 0.5 mM NADPH for 45 min before an addition of 10% formic acid to quench the reactions. The ratio of γ-Cl-**4**-$ACP_{II}$ and γ-Cl-**6**-$ACP_{II}$ were measured by IRMPD-based quantification as described above. (Performed by Bo Wang and Liangcai Gu)

**Modeling the PPant arm of (4*S*)-γ-Cl-2-$ACP_{II}$ into Cur $ECH_2$ structure.** Initial atomic coordinates and topology files for the PPant arm of (4*S*)-γ-Cl-**2**-$ACP_{II}$ were generated using the PRODRG2 server (*51*). The PPant arm was initially positioned into Cur $ECH_2$ (PDB code 2Q34) as previously described (*34*). Any overlapping waters were removed before energy minimization after the addition of hydrogens using the program CNS 1.2 (*52*) in 1080 steps of conjugate gradient minimization using no experimental energy terms and full harmonic restraints. (Performed by Todd Geders)

## 4.6 SUPPLEMENTAR FIGURES AND TABLES

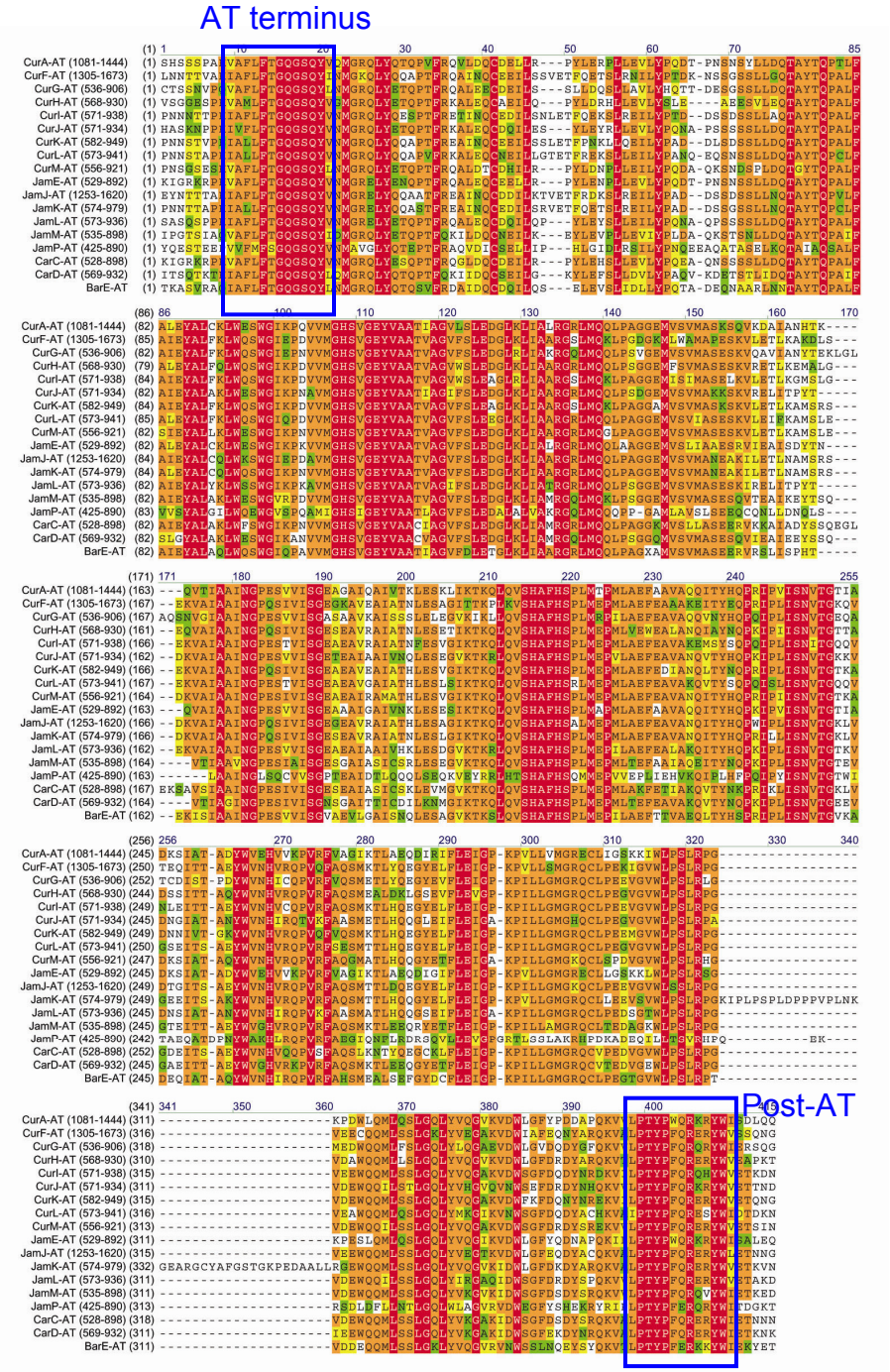



**Figure 4-S1. Alignment of AT domains of the sequenced pathways from *L. majuscula*, including curacin, jamaicamide, barbamide and putative carmabin pathways.**

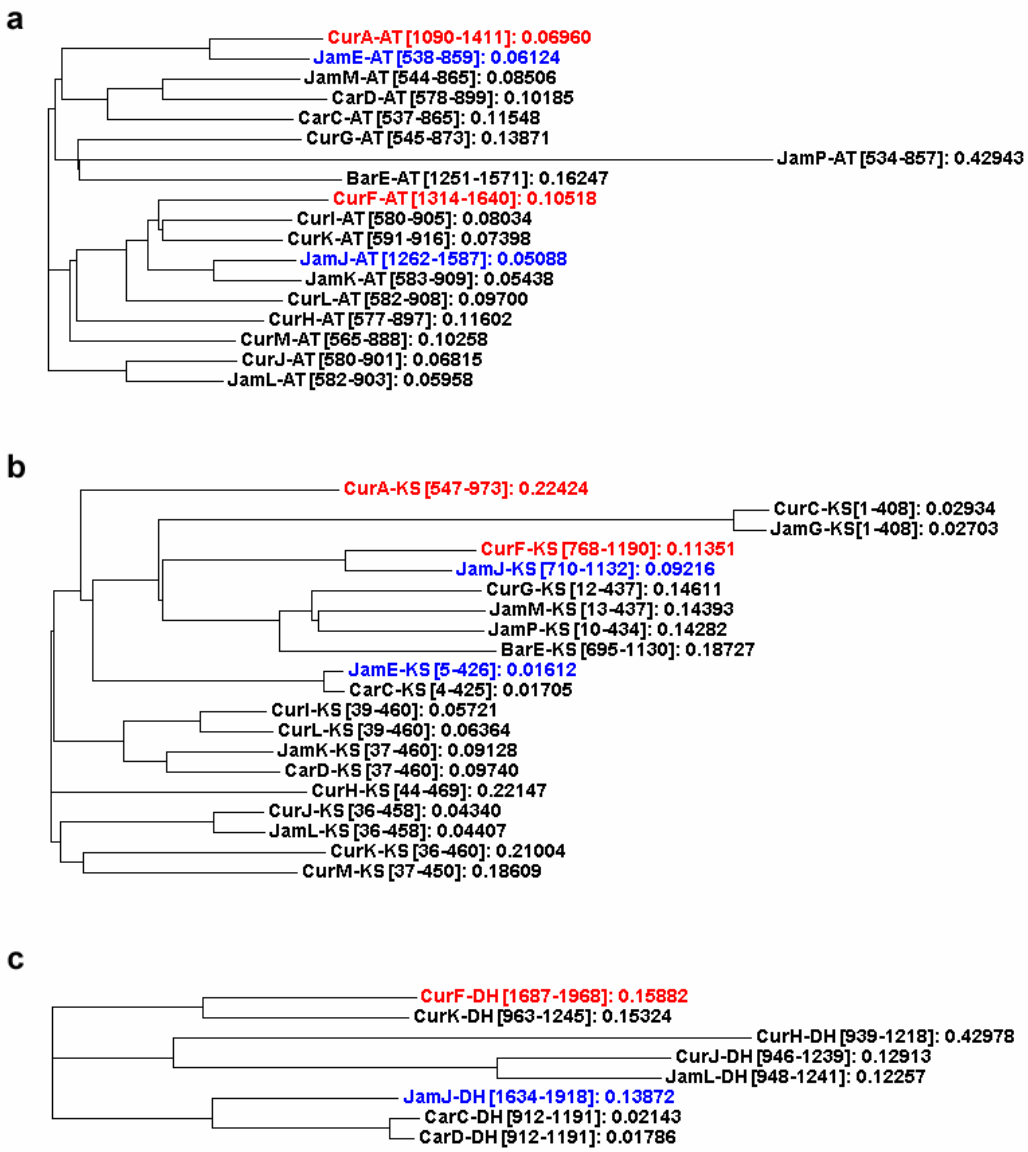



**Figure 4-S2. Phylogenetic trees for a, AT, b, KS and c, DH domain sequences of curacin, jamaicamide, barbamide and putative carmabin pathways from *L. majuscula*.** The phylogram trees with distances were constructed by ClustalW2.

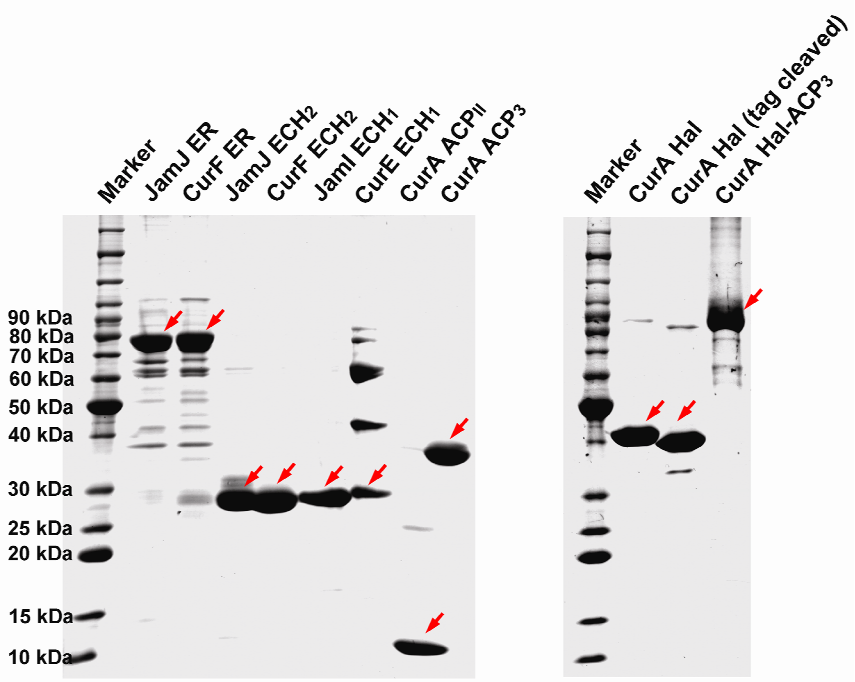



**Figure 4-S3. Coomassie blue-stained SDS-PAGE of purified Cur and Jam enzymes.** JamJ ER and CurF ER were expressed as N-terminal GST-tagged proteins to increase their solubility. CurE $ECH_1$ was coexpressed with chaperones to increase its solubility and coeluted with them from the Ni-affinity column. His-tag cleaved CurA Hal was prepared for metal content analysis. The target enzymes are indicated by red arrows.

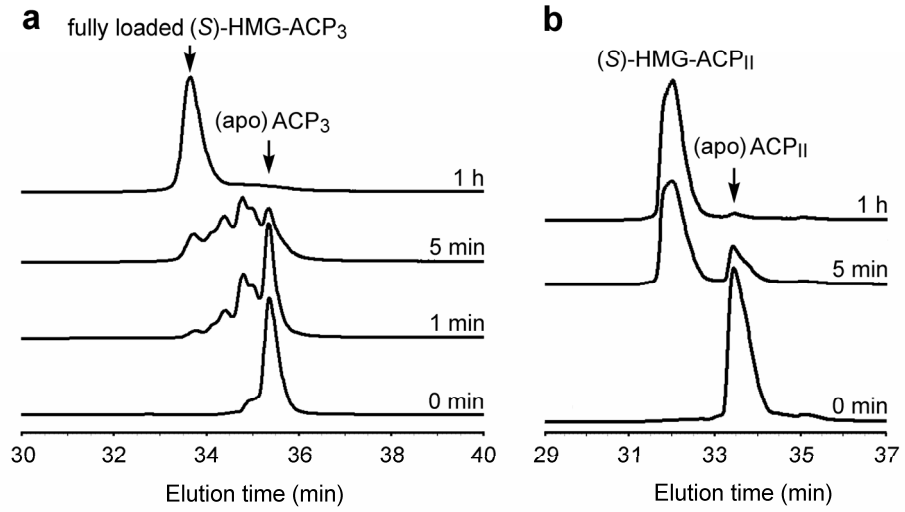



**Figure 4-S4. HPLC analysis of HMG substrate loading onto a, (apo) $ACP_3$, b, (apo) $ACP_{II}$.** Multiple peaks were observed for (apo) $ACP_3$ substrate loading due to mono-, di- and tri-loading with (*S*)-HMG-CoA.

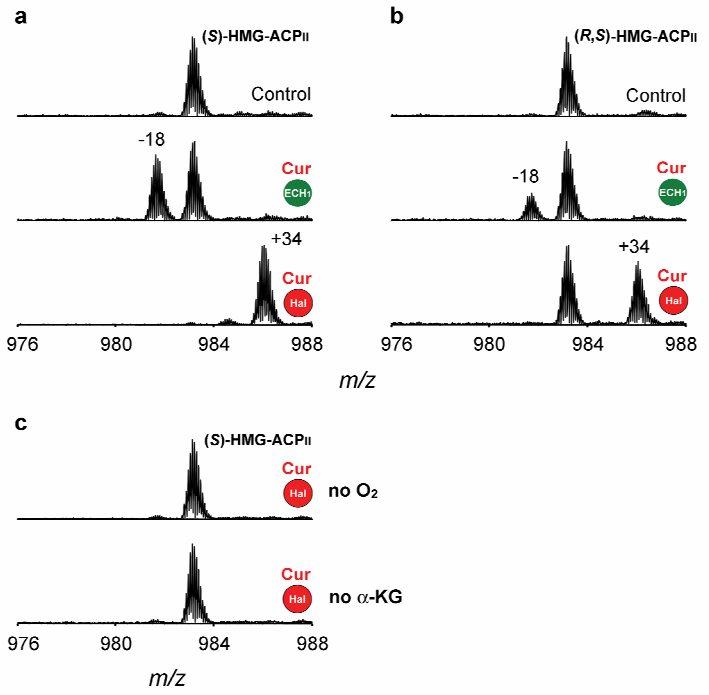



**Figure 4-S5. a, b, HMG substrate chirality preference of Cur Hal and $ECH_1$; c, $O_2$ and α-KG dependence of Cur Hal.** Partial FTICR mass spectra for $ECH_1$ and Hal reactions with **a**, (*S*)-HMG-$ACP_{II}$ substrate, and **b**, (*R*,*S*)-HMG-$ACP_{II}$ substrate. **c**, Partial FTICR mass spectra for Hal reactions without $O_2$ or α-KG. For Hal reaction without $O_2$, the reaction mixture was prepared in glove box and capped during its incubation.

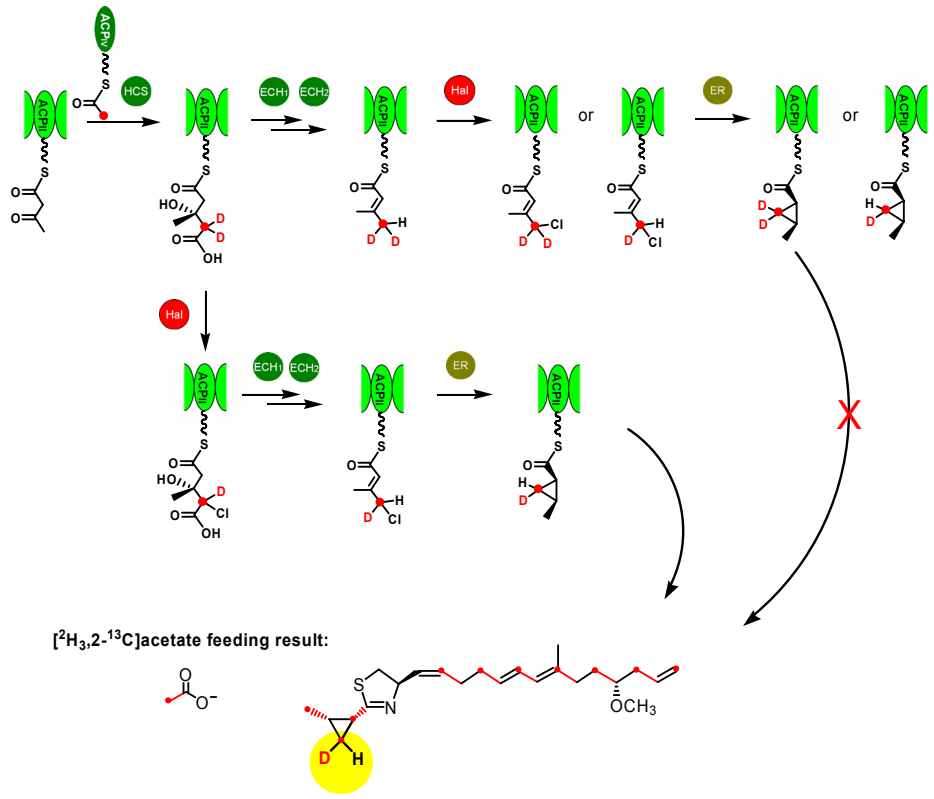



**Figure 4-S6. Feeding experiment to demonstrate that Hal chlorination precedes $ECH_2$ decarboxylation.** The β-branching carbon forming the cyclopropane was labeled with only one deuterium atom by [$^2H_3$,2-$^{13}$C]acetate supplied in the growth media of *L. majuscula*. If Hal chlorination happens after $ECH_2$ decarboxylation, the β-branching carbon would be labeled randomly with one or two deuterium atoms.

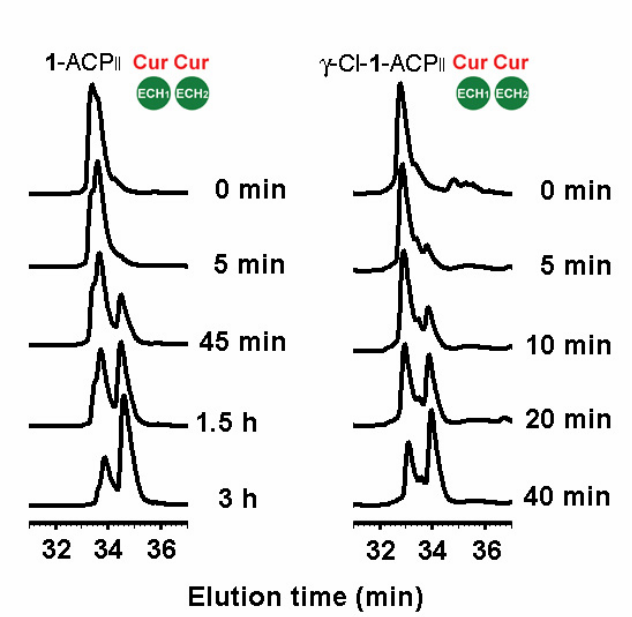



**Figure 4-S7. HPLC analysis of $ECH_1/ECH_2$ coupled dehydration and decarboxylation to compare the rates for the chlorinated and non-chlorinated substrate.** The reactions were performed under a uniform reaction condition, and quenched after the indicated periods of time.

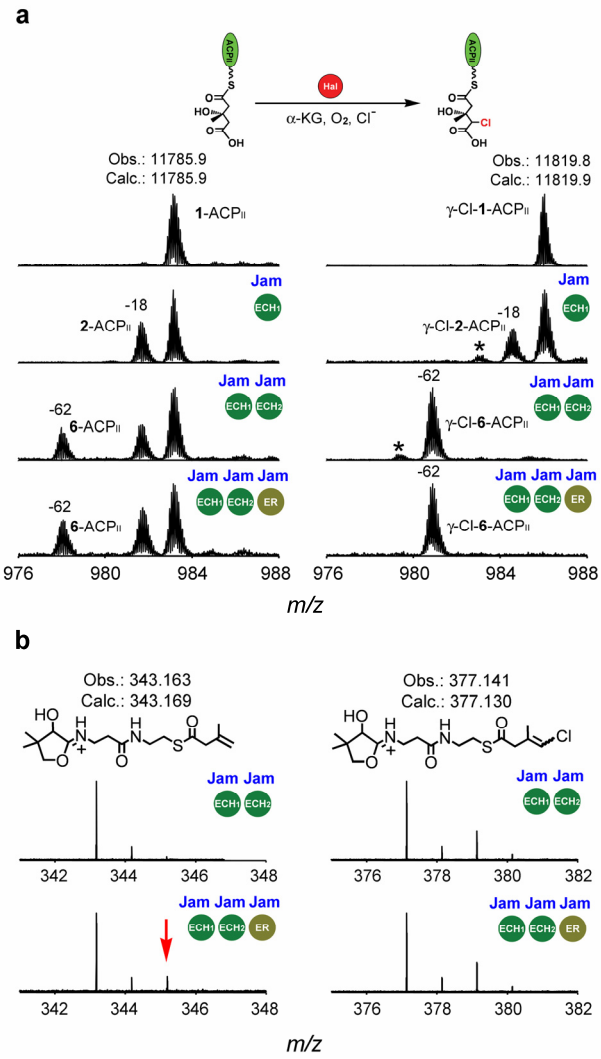



**Figure 4-S8. FTICR mass spectra and IRMPD spectra for Jam enzyme reactions**. **a**, Partial FTICR mass spectra for Jam $ECH_1$, $ECH_2$ and ER reactions. **1**-$ACP_{II}$ was chlorinated by Cur Hal for 1 h to generate γ-Cl-**1**-$ACP_{II}$ substrate. The reactions were incubated at 30°C for 2 h for the **1**-$ACP_{II}$ substrate and 30 min for the γ-Cl-**1**-$ACP_{II}$ substrate. **b**, IRMPD spectra for the PEPs of **6**-$ACP_{II}$ and γ-Cl-**6**-$ACP_{II}$ before Jam ER saturation (above) and after Jam ER saturation (below). A small amount of saturated product with its PEP indicated by red arrow was observed to be mixed with the **6**-$ACP_{II}$. Based on the butylamine cleavage analysis, **3**-$ACP_{II}$ was inferred to be the minor product of Jam $ECH_2$ when using the non-chlorinated substrate.

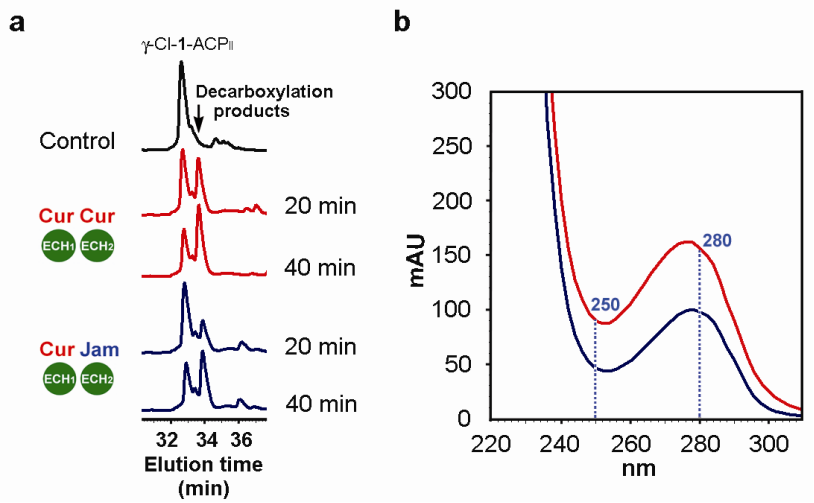



**Figure 4-S9. The UV absorption difference for Cur and Jam $ECH_2$ decarboxylation products. a**, HPLC traces showing the separation of γ-Cl-**1**-$ACP_{II}$ and its decarboxylation products catalyzed by Cur and Jam $ECH_1$/$ECH_2$. **b**, UV spectra (220-310 nm) of Cur (red) and Jam (blue) $ECH_2$ decarboxylation products.

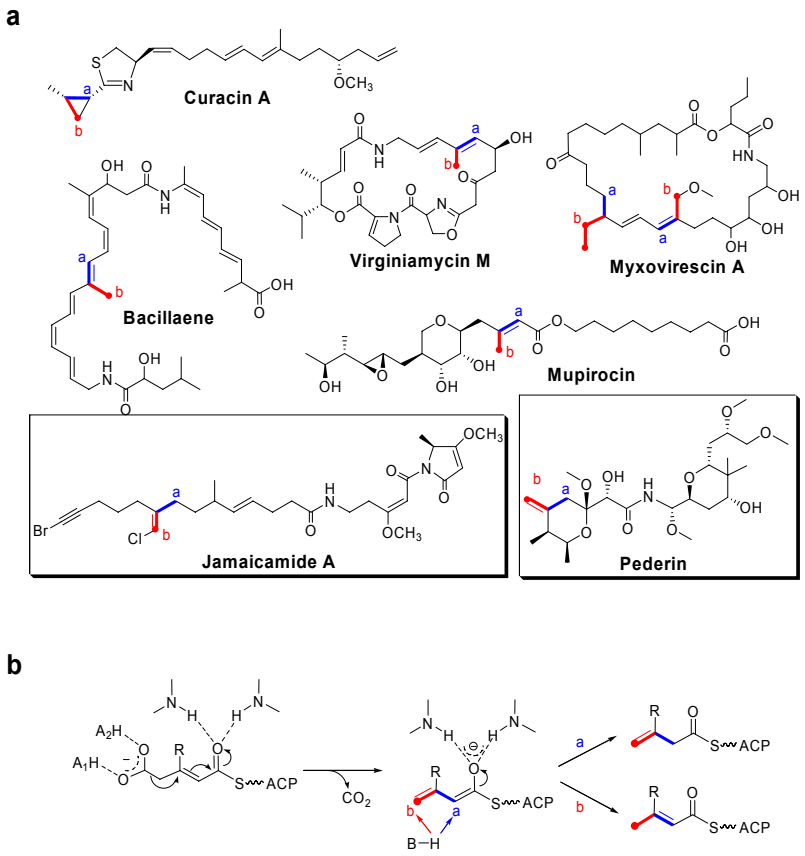



**Figure 4-S10. Natural products with different β-branching-associated C=C positions via $ECH_2$ regiochemical control. a**, Two groups of natural products with the α,β and β,γ C=Cs adjacent to their β-branching carbons. The C=Cs of curacin A and myxovirescin A have been modified by ERs. **b**, The mechanism for the formation of α,β and β,γ C=C products.

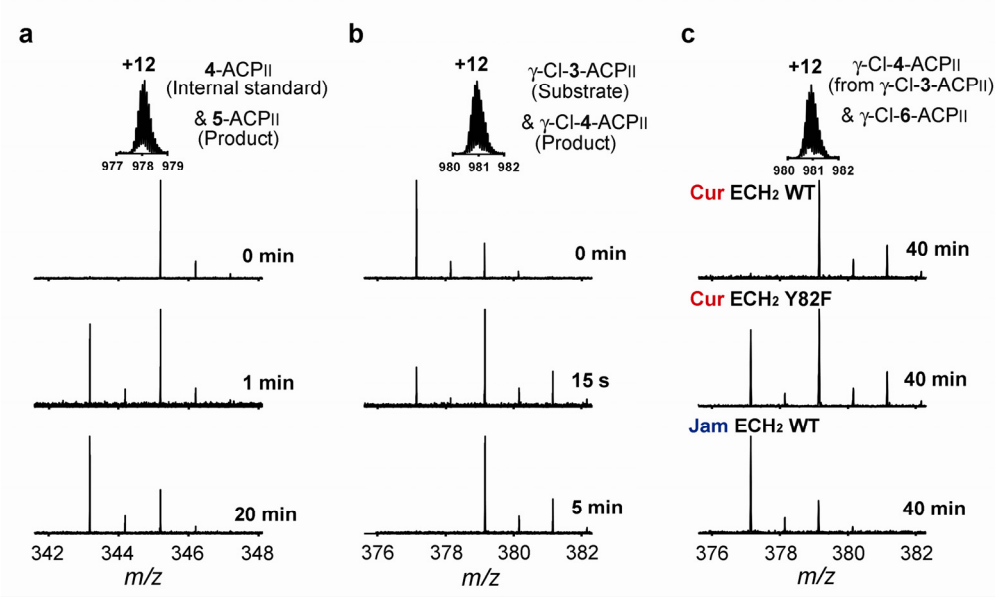



**Figure 4-S11. IRMPD-based quantification to measure the yields of a, Cur ER cyclopropanation, b, Jam ER saturation, and c, the ratio of α,β and β,γ C=C products of $ECH_2$s.** Because the $ACP_{II}$ species with two Dalton mass difference are equally distributed over different charge states, we focused on the +12 change state to record IRMPD spectra. **a**, To measure the yield of Cur ER cyclopropanation, **4**-$ACP_{II}$ was added as internal standard for **5**-$ACP_{II}$. **b**, The yield of Jam ER saturation was measured by the ratio of γ-Cl-**3**-$ACP_{II}$ substrate and γ-Cl-**4**-$ACP_{II}$ product. **c**, To measure the ratio of α,β and β,γ C=C products, the α,β C=C product was reduced by Jam ER by treating the mixtures of α,β and β,γ C=C products with Jam ER for 30 min before IRMPD analysis.

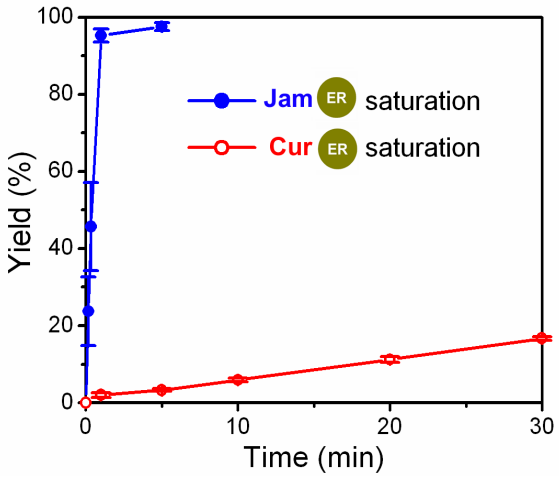



**Figure 4-S12. Comparison of Cur and Jam ER saturation by using 3-ACP$_{II}$ substrate.** IRMPD-based quantification was applied to measure the yields of ER saturation.

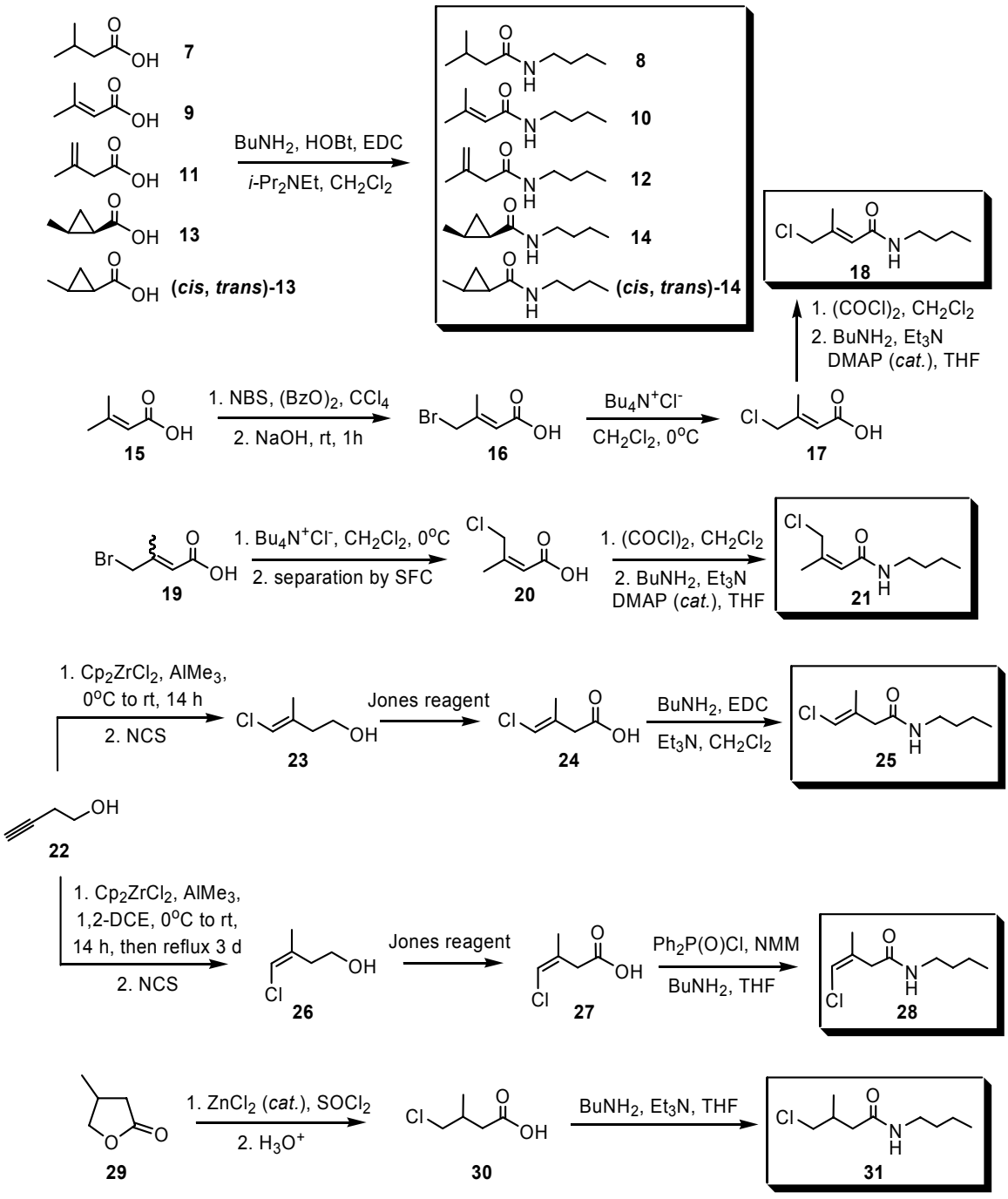



**Figure 4-S13**. **Synthesis of butylamide derivatives as GC-MS authentic standards.**

**Table 4-S1. Primers for expression constructs.**

| Primer name | Sequences (5' to 3') |
|---|---|
| CurA-Hal-for | CATATGAACCGGGAACAAGTTGAACAA (*Nde*I) |
| CurA-Hal-rev | CTCGAGTTACGAAGAGATGCTTGGTGTTTC (*Xho*I) |
| CurA-$ACP_3$-rev | CTCGAGCTCCTAAACTGTAACCTGTTT (*Xho*I ) |
| CurF-ER-for | CCGCGGGTATACACCAGGTAACCTTTAAC (*Sac*II) |
| CurF-ER-rev | CTCGAGATTTTTCTTTGTTGGGGA (*Xho*I) |
| JamI-$ECH_1$-for | CATATGTATTACCAAACCCTAAAA (*Nde*I) |
| JamI-$ECH_1$-rev | CTCGAGGCTTTGCCATGGATATAAC (*Xho*I) |
| JamJ-$ECH_2$-for | CATATGGCAAAGCTGAACTTGAATC (*Nde*I) |
| JamJ-$ECH_2$-rev | CTCGAGCTGCTGGAAGGTTTTTTC (*Xho*I) |
| JamJ-ER-for | CCGCGGGTATACATCAAACCACTAGTCAA (*Sac*II) |
| JamJ-ER-rev | CTCGAGTTCTTTTGTCGATTCTGGTTC (*Xho*I) |

The restriction sites are underlined.

**Table 4-S2. ESI-FTICR-MS analysis.**

| $ACP_{II}$ samples | ESI-FTICR-MS | | IRMPD (PEP) | |
|---|---|---|---|---|
| | Obs. avg mass | Calc. avg mass | Obs. $[M + H]^+$ | Calc. $[M + H]^+$ |
| apo-$ACP_{II}$ | 11300.7 | 11300.8 | | |
| **1**-$ACP_{II}$ | 11785.9 | 11785.9 | 387.164 | 387.159 |
| **2**-$ACP_{II}$ | 11767.9 | 11767.9 | 405.174 | 405.170 |
| **3**-$ACP_{II}$ | 11723.9 | 11723.9 | 343.174 | 343.169 |
| **6**-$ACP_{II}$ | 11723.9 | 11723.9 | 343.163 | 343.169 |
| **4**-$ACP_{II}$ | 11725.9 | 11725.9 | 345.190 | 345.185 |
| γ-Cl-**1**-$ACP_{II}$ | 11819.8 | 11819.9 | 439.128 | 439.131 |
| γ-Cl-**2**-$ACP_{II}$ | 11802.0 | 11801.9 | 421.110 | 421.120 |
| γ-Cl-**3**-$ACP_{II}$ | 11757.8 | 11757.9 | 377.140 | 377.130 |
| γ-Cl-**6**-$ACP_{II}$ | 11757.8 | 11757.9 | 377.141 | 377.130 |
| γ-Cl-**4**-$ACP_{II}$ | 11759.8 | 11759.9 | 379.156 | 379.146 |
| **5**-$ACP_{II}$ | 11723.9 | 11723.9 | 343.169 | 343.169 |

Masses are in Daltons. Calc., calculated; Obs., observed.
PEP, PPant ejection product.

## 4.7 REFERENCES


1. T. Hartmann, *Entomol. Exp. Appl.* **80**, 177 (1996).
2. R. D. Firn, C. G. Jones, *Nat. Prod. Rep.* **20**, 382 (2003).
3. H. Jenke-Kodama, R. Mueller, E. Dittmann, *Prog. Drug Res.* 119 (2008).
4. M. A. Fischbach, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4601 (2008).
5. E. Haslam, *Chemoecology* **5**, 89 (1994).
6. F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. Garneau-Tsodikova, C. T. Walsh, *Chem. Rev.* **106**, 3364 (2006).
7. M. A. Fischbach, J. Clardy, *Nat. Chem. Biol.* **3**, 353 (2007).
8. E. Haslam, *Nat. Prod. Rep.* **3**, 217 (1986).
9. C. Khosla, R. S. Gokhale, J. R. Jacobsen, D. E. Cane, *Annu. Rev. Biochem.* **68**, 219 (1999).
10. B. S. Moore, C. Hertweck, *Nat. Prod. Rep.* **19**, 70 (2002).
11. M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **106**, 3468 (2006).
12. L. C. Gu *et al.*, *J. Am. Chem. Soc.* **128**, 9014 (2006).
13. C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh, P. C. Dorrestein, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8977 (2006).
14. D. J. Edwards *et al.*, *Chem. Biol.* **11**, 817 (2004).
15. P. Verdier-Pinard *et al.*, *Mol. Pharmacol.* **53**, 62 (1998).
16. Z. X. Chang *et al.*, *J. Nat. Prod.* **67**, 1356 (2004).
17. M. Costas, M. P. Mehn, M. P. Jensen, L. Que, *Chem. Rev.* **104**, 939 (2004).
18. L. C. Blasiak, F. H. Vaillancourt, C. T. Walsh, C. L. Drennan, *Nature* **440**, 368 (2006).
19. F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. E. O'Connor, C. T. Walsh, *Nature* **436**, 1191 (2005).
20. F. H. Vaillancourt, J. Yin, C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10111 (2005).
21. D. P. Galonic, F. H. Vaillancourt, C. T. Walsh, *J. Am. Chem. Soc.* **128**, 3900 (2006).
22. Z. X. Chang *et al.*, *Gene* **296**, 235 (2002).
23. P. M. Flatt *et al.*, *J. Nat. Prod.* **69**, 938 (2006).
24. D. P. Galonic, E. W. Barr, C. T. Walsh, J. M. Bollinger, C. Krebs, *Nat. Chem. Biol.* **3**, 113 (2007).
25. C. Krebs, D. G. Fujimori, C. T. Walsh, J. M. Bollinger, *Acc. Chem. Res.* **40**, 484 (2007).
26. C. T. Calderone, D. F. Iwig, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, *Chem. Biol.* **14**, 835 (2007).

27. H. Jiang *et al.*, *J. Am. Chem. Soc.* **130**, 6336 (2008).
28. A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson, C. M. Thomas, *J.Biol. Chem.* **280**, 6399 (2005).
29. E. Nordling, H. Jornvall, B. Persson, *Eur. J. Biochem.* **269**, 4267 (2002).
30. V. Simunovic *et al.*, *Chembiochem* **7**, 1206 (Aug, 2006).
31. G. L. Tang, Y. Q. Cheng, B. Shen, *Chem. Biol.* **11**, 33 (2004).
32. S. Sudek *et al.*, *J. Nat. Prod.* **70**, 67 (2007).
33. J. Piel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16222 (2004).
34. T. W. Geders *et al.*, *J. Biol. Chem.* **282**, 35954 (2007).
35. Y. Y. Tang, C. Y. Kim, Mathews, II, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11124 (2006).
36. H. Jenke-Kodama, T. Borner, E. Dittmann, *PLoS Comput. Biol.* **2**, 1210 (2006).
37. C. P. Ridley, H. Y. Lee, C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4595 (2008).
38. P. C. Dorrestein *et al.*, *Biochemistry* **45**, 12756 (Oct 24, 2006).
39. J. Kopka, J. B. Ohlrogge, J. G. Jaworski, *Anal. Biochem.* **224**, 51 (1995).
40. R. A. Butcher *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1506 (2007).
41. A. K. El-Sayed *et al.*, *Chem. Biol.* **10**, 419 (2003).
42. N. Pulsawat, S. Kitani, T. Nihira, *Gene* **393**, 31 (2007).
43. J. Piel, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14002 (2002).
44. A. V. Ramaswamy *et al.*, in *Front. Mar. Biotechnol.* (2006), pp. 175-224.
45. K. M. Bischoff, V. W. Rodwell, *Biochem. Med. Metab. Biol.* **48**, 149 (1992).
46. F. H. Vaillancourt, S. Han, P. D. Fortin, J. T. Bolin, L. D. Eltis, *J. Biol. Chem.* **273**, 34887 (1998).
47. L. C. Gu *et al.*, *Science* **318**, 970 (2007).
48. M. W. Senko, J. D. Canterbury, S. H. Guan, A. G. Marshall, *Rapid Commun. Mass Spectrom.* **10**, 1839 (1996).
49. E. B. Ledford, D. L. Rempel, M. L. Gross, *Anal. Chem.* **56**, 2744 (1984).
50. E. F. Gordon, B. A. Mansoori, C. F. Carroll, D. C. Muddiman, *J. Mass Spectrom.* **34**, 1055 (1999).
51. A. W. Schuttelkopf, D. M. F. van Aalten, *Acta Crystallogr.D.* **60**, 1355 (2004).
52. A. T. Brunger *et al.*, *Acta Crystallogr.D.* **54**, 905 (1998).

**Notes:**

Liangcai Gu and David Sherman designed and performed the experiments;

Bo Wang and Kristina Håkansson recorded and analyzed the FT-ICR MS data;
Amol Kulkarni and Peter Wipf synthesized chlorinated standards for GC-MS;
Todd Geders and Janet Smith designed the site-directed mutagenesis.

# CHAPTER 5
# DECARBOXYLATIVE CHAIN TERMINATION

## 5.1 SUMMARY

In this chapter, a novel chain termination module that mediates a decarboxylative chain termination leading to the formation of terminal olefin in the curacin A pathway is biochemically assessed. First, we identified the correct *curM* sequence by sequencing the thioesterase (TE) region in the selected cosmids from the genomic library of *Lyngbya majuscula*. Next, the acyl carrier protein (ACP), sulfotransferase (ST) and thioesterase domains from CurM chain extension and termination modules were overexpressed, purified and biochemically assessed. The ST domain was demonstrated to transfer a sulfuryl group from adenosine 3′-phosphate 5′-phosphosulfate (PAPS) to the β-hydroxyl group of the model substrates. The TE domain was proved to release the acyl chain by cleaving the thioester bond. New model substrates were designed to facilitate the detection of the terminal olefin products in ST/TE coupled assays. This ST-TE chain termination model can serve as an important engineering tool to generate a hydrophobic end in polyketide products to improve their physicochemical properties.

## 5.2 INTRODUCTION

Modular polyketide synthases (PKSs) are giant biosynthetic machines that catalyze programmed metabolic pathways to assemble short-chain acyl-CoA building blocks into complex polyketide products by one step of chain initiation followed by multiple steps of chain elongation and one step of chain releasing. So far, three major strategies have been reported for cleavage of the thioester bond between the thiol group of ACP/peptidyl carrier protein (PCP) PPant arm and PKS/non-ribosomal peptide synthase (NRPS) acyl/peptidyl chain after the final step of chain extension completes (*1*). The first strategy

is to hydrolyze the full-length acyl, peptidyl, or hybrid chain from carrier protein PPant thiol to release the free acid via formation of an acyl-*O*-TE intermediate and releasing the acyl group by nucleophilic attack with $H_2O$ (*1*). The second strategy and the one of most medicinal interest is to use an intramolecular nucleophile within the acyl chain to release the acyl group from the acyl-*O*-TE intermediate (*1*). This intramolecular nucleophile can be N-terminal or side chain -$NH_2$s that give cyclic lactams, or side chain -OHs that result in macrolactones. The last strategy of chain release is to employ a reductase (Red) domain rather than the TE domain that utilizes NAD(P)H as a hydride-transfer cosubstrate (*1*). Usually, the most downstream catalytic domain in such assembly lines is a reductase domain instead of a TE domain. The two-electron reduction of an acyl thioester generates the thioacetal, which spotaneously decomposes to release an aldehyde product. In some cases, the aldehyde can undergo a second round of reduction to give an alcohol.

Curacin A is a potent antimitotic agent isolated from *L. majuscula* (*2, 3*). As a mixed polyketide/non-ribosomal peptide, it has unusually high hydrophobicity. All the keto groups from the chain extension steps are reduced to hydroxyl groups, which are removed by the dehydratase (DH)-catalyzed dehydration or modified by O-methyltransferase (OM)-catalyzed methylation. More interestingly, the curacin A has an unusual terminal olefin instead of a carboxyl, aldehyde or alcohol group. The identification of curacin A biosynthetic gene cluster (*3*) revealed that the terminal olefin group is most likely generated from the decarboxylation/dehydration catalyzed by an unprecedented ST-TE chain termination module. Based on the feeding study and the pathway annotation (*3*), a TE and ST-mediated decarboxylative chain termination mechanism was proposed (Figure 5-1).

Sulfotransferases are enzymes that catalyze the transfer of a sulfuryl group from the donor, PAPS, to the receptors ranging from proteins to compounds (*4*). Two classes of STs have been identified and characterized: cytosolic STs and membrane-associated STs. Cytosolic STs sulfonate small endogenous and exogenous compounds, such as hormones, bioamines, drugs, and various xenobiotic compounds, which are involved in detoxification, hormone regulation, and drug metabolism (*4*). Membrane-associated STs,

many of which have been implicated recently in crucial biological processes, sulfonate

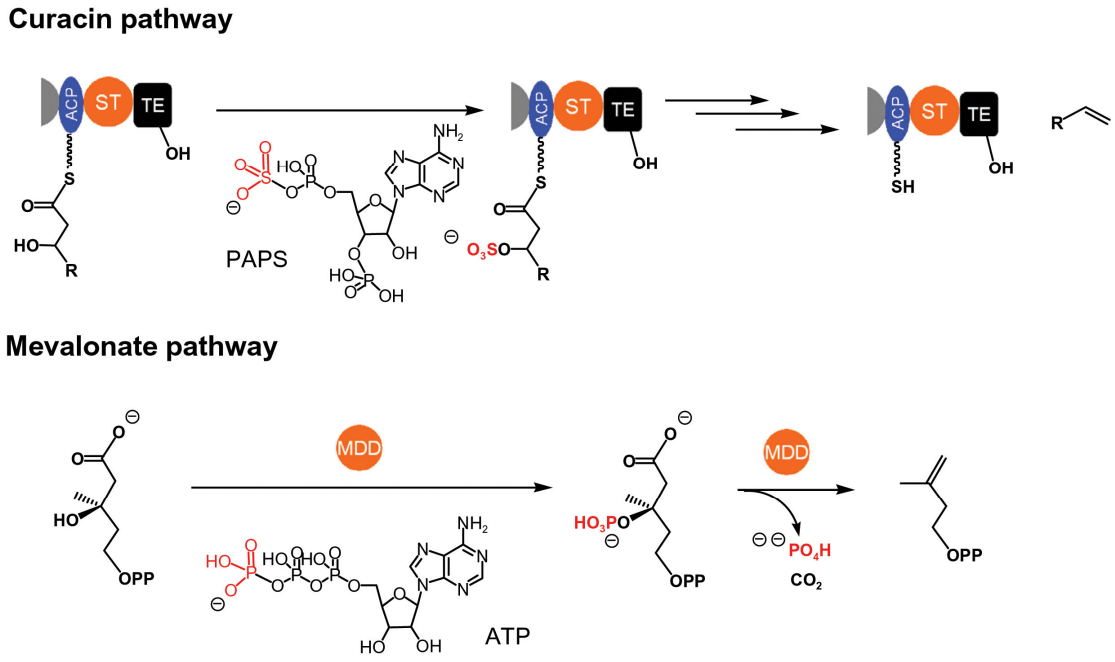



**Figure 5-1. The terminal olefin formation via decarboxylative elimination mechanisms. A.** MDD-catalyzed decarboxylative elimination. **B.** Proposed chain termination mechanism in curacin A pathway.

larger biomolecules, such as carbohydrates and proteins, and play key roles in a number of molecular-recognition events and biochemical signaling pathways (*4*). Here, CurM ST is unprecedented in that it is embedded in PKS modules and recruited for polyketide chain termination process.

The sulfotransferase superfamily members have a conserved structural fold, a core of four to five parallel beta-strands flanked by alpha helices. The PAPS binding region is highly conserved both in primary and tertiary structures (*4*). Given the diverse nature of sulfuryl group receptor, it is expected that the substrate-binding region has the most variation (*4*). Mechanistic insights have been derived from the solved X-ray structures of some STs and their biochemical studies. It was demonstrated that the sulfonation reactions proceed by an in-line attack of the nucleophile at the sulfate group of PAPS. However, it is yet to be confirmed that whether the conserved active-site histidine residue assists in the reaction by deprotonating the substrate hydroxyl group, or acts as a nucleophile to form an unstable protein-sulfate complex (*5*).

This unusual decarboxylative chain termination chemistry provides us an useful tool to generate a terminal olefin into polyketide-related compounds. Presumably, the ST-catalyzed sulfonation of β-hydroxyl group can promote the decarboxylative elimination by recruiting the sulfate as an excellent leaving group. However, the timing of sulfonation and TE cleavage, as well as the enzyme that mediates the decarboxylative elimination is not evident. Notably, similar strategy to generate the terminal olefin was observed in the mevalonate pathway in isoprenoid biosynthesis. A kinase, mevalonate-5-diphosphate decarboxylase (MDD), was proved to convert β-hydroxyl group in mevalonate-5-diphosphate to phosphate group, which also serves as a leaving group, and catalyze successive decarboxylative elimination leading to a isopentenyl diphosphate product (Figure 5-1) (*6*). However, a similar terminal olefin in tautomycetin (TMC), a polyketide metabolite isolated from *Streptomyces* sp. CK4412 and identified as an immunosuppressor of activated T cells, is likely generated by a distinct unknown mechanism (*7*). In tautomycetin biosynthetic pathway, only TE was observed in the chain termination module, and some tailoring enzymes, such as a P450 and decarboxylase, were suggested to mediate the terminal olefin formation (*7*).

The ST-TE chain termination module provides us a useful bioengineering tool to generate a highly hydrophobic functional group in the polyketide-related compounds, which might optimize the physicochemical properties of some natural products.

## 5.3 RESULTS

### 5.3.1 Resequencing of CurM TE Domain

Based on the bioinformatic analysis and protein expression results, the previously published sequence of the end of curacin A pathway that starts from the middle of CurM TE region (*3*) was predicted to be incorrect. Only a part of thioesterase domain was found in this TE region based on the Conserved Domain (CD) searching and sequence alignment. No soluble protein expression was achieved for any CurM TE-containing constructs, such as ACP-ST-TE, ST-TE and TE.

We sought to identify the cosmids that contain the correct curacin A downstream pathway. Thus, the selected cosmids from the genomic library (*8*) that contain PKS or NRPS modules were screened by PCR using a primer pair that amplifies *curM* ST gene. The pLM19 and pLM14 cosmids (*3*) were detected to contain *curM* ST gene. These two cosmids were further sequenced by the primer walking from CurM TE to ~300 base pair downstream. Our results confirmed that the published sequence (*3*) of CurM TE and its downstream regions in pLM17 does not belong to the curacin A pathway. This finding is supported by the result from the *L. majuscula* single cell genome sequencing project provided by William Gerwick lab at UCSD (Figure 5-2).

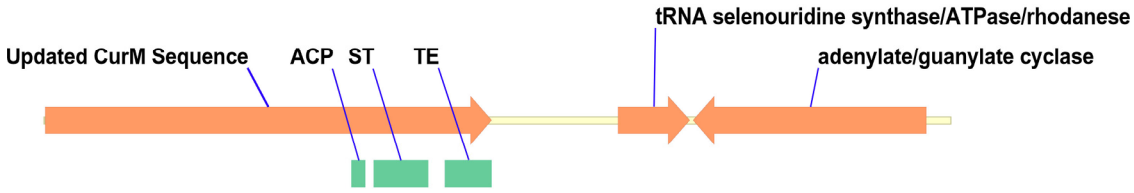



**Figure 5-2. The updated curacin A downstream pathway from *L. majuscular* genome sequencing project**. (Gerwick lab)

Based on the corrected sequence of curacin A downstream pathway, this pathway ends at *curM*, and no *curN* (*3*) was found in the downstream region. The new CurM TE sequence was predicted to contain a full conserved region for thioesterases.

### *5.3.2 CurM ST Assays*

CurM ST was biochemically assessed to identify and characterize its activities. The ST was produced as an N-terminal His-tagged protein (Figure 5-3). Two long chain (C14) model substrates with (*R*) or (*S*) β-hydroxyl and δ-methoxyl groups were synthesized as the CoA and SNAC thioesters. The CoA substrates were also loaded onto (apo) CurM ACP to generated ACP-linked substrates. All CoA, SNAC and ACP-linked substrates were proved to be sulfonated by CurM ST. Thus, we focused on the substrates in CoA thioester form to examine the stereoselectivity for the β-hydroxyl group. The CurM ST transfers sulfuryl group from PAPS to both substrates with (*R*) and (*S*) β-hydroxyl groups (Figure 5-4).

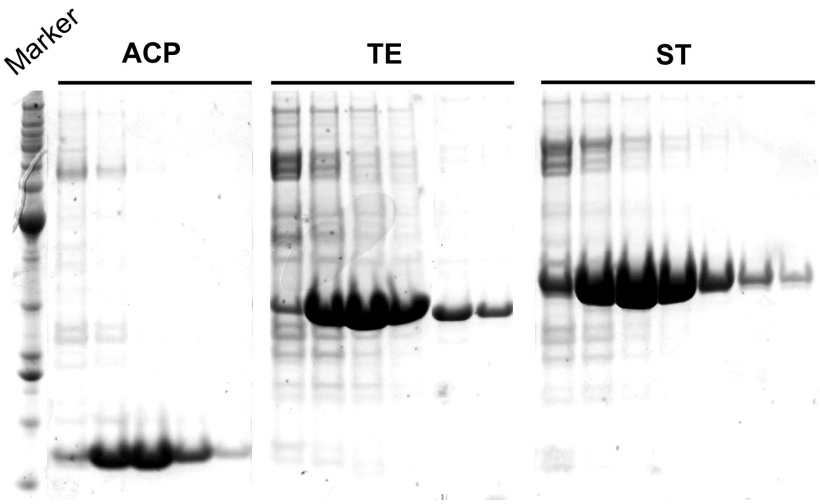



**Figure 5-3. Ni-NTA purifications of CurM ACP, TE and ST.**

However, more sulfonation product of the (*R*) β-hydroxyl group substrate accumulated than the (*S*) β-hydroxyl group substrate. Interestingly, a considerable amount of α, β dehydration product was observed for the (*S*) β-hydroxyl group substrate, which suggests that (*S*) β-sulfate group might serve as a leaving group for ST-catalyzed dehydration.

### *5.3.3 CurM TE Assays*

The CurM TE was produced as a partially soluble C-terminal His-tagged protein (Figure 5-3). The model substrates in the CoA thioester form were examined for the TE stereoselectivity of the β-hydroxyl group. The CurM TE was proved to cleave the thioester bond of both model substrates with the (*R*) and (*S*) β-hydroxyl groups (Figure 5-4). However, according to our preliminary HPLC

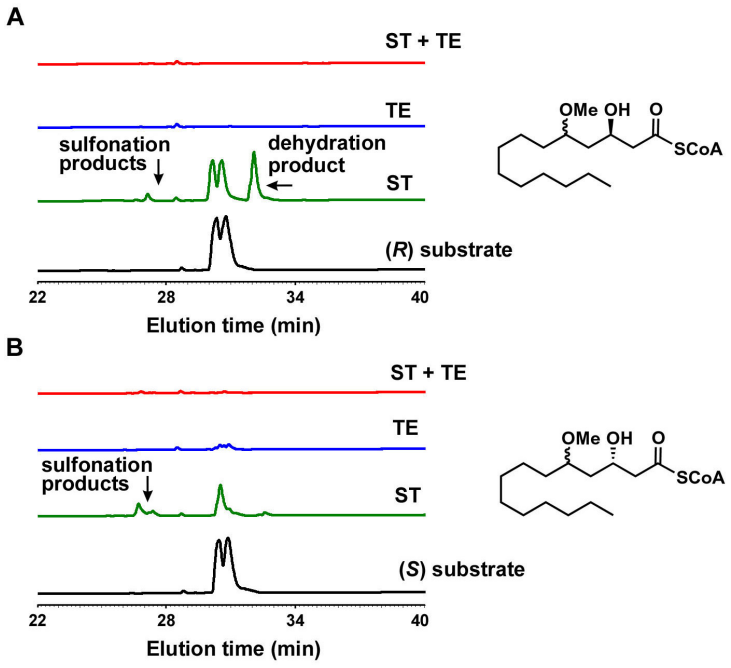



**Figure 5-4. HPLC analysis for the ST and TE assays.**

analysis, the CurM TE shows a higher catalytic acitivity for the (*R*) β-hydroxyl group substrate than (*S*) β-hydroxyl group substrate, suggesting that the native substrate of CurM ST should contain a (*R*) β-hydroxyl group.

## 5.4 FUTURE DIRECTIONS

The terminal olefin products of the chain model substrates can only be detected by GC-MS, which is not facile method for the follow-up kinetic assays. Thus, we decided to introduce a hydrophilic imidazole group as a chromophore into the model substrates (Figure 5-5) to facilitate the HPLC analysis. These substrates will be synthesized by Peter Wipf group at University of Pittsburgh.

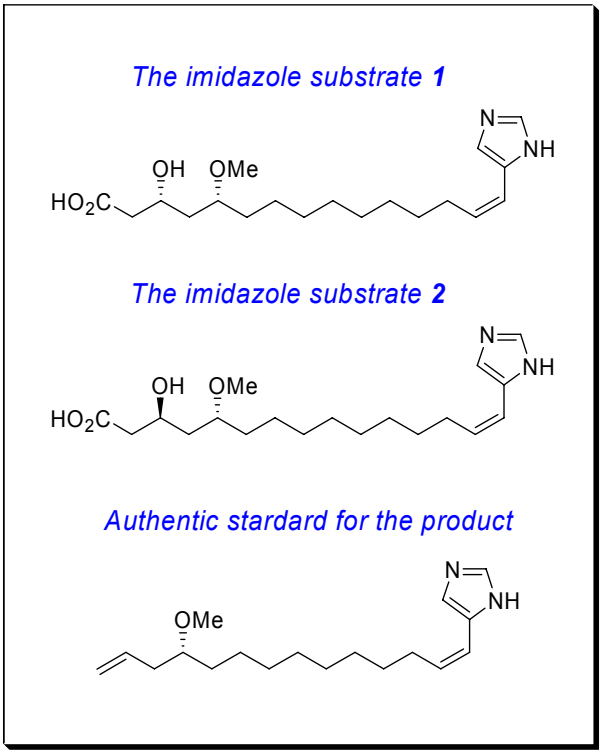



**Figure 5-5. Imidazole-containing model substrates and the authentic standard for the HPLC analysis of terminal olefin products.**

Given our results from the CurM ST and TE assays, some questions are yet to be answered: (1) the timing of ST-catalyzed sulfonation. Does it happen before or after the TE cleavage? We need to test whether CurM ST can catalyze the sulfonation of the carboxylic acids after the TE cleavage. (2) Which enzyme catalyzes the decarboxylative elimination, ST or TE? The ST-catalyzed dehydration of the CoA thioesters implies that the activity of ST might not be limited to the β-hydroxyl sulfonation. The new designed imidazole-containing model substrates will enable us to trace their intermediates and products in ST/TE-coupled assays.

## 5.5 MATERIAL AND METHODS

**Chemicals.** PAPS was from Sigma-Aldrich (>60% purity). The model substrates in carboxyl acid forms were synthesized by Peter Wipf group at University of Pittsburgh. The CoA or SNAC thioesters of the model substrates were synthesized as described in chapter 4. All other chemicals were from Sigma-Aldrich.

**Bacterial strains, media and culture conditions.** *Escherichia coli* DH5α MCR (Invitrogen) was used for DNA propagation. *Escherichia coli* BL21 (DE3) transformed with the derivatives of pET24b and pET28b were used for protein overexpression in Luria-Bertani (LB) medium. Ampicillin (100 μg/mL), carbenecillin (100 μg/mL), kanamycin (50 μg/mL), and apramycin (50 μg/mL) were used for the corresponding plasmid construct resistance marker selection in *E. coli* cultures.

**Resequencing of *curM*.** The selected cosmids from the cosmid library (*8*) of *L. majuscula* were screened by PCR using a pair of primers to amplify CurM ST gene: (F) GGA TGC GGA TGC AAA AAC TTG, and (R) CGG ATG CAA AAA CTT GTC GGG. According to the PCR results, pLM19 and pLM14, as well as pLM17 (*8*) were detected to contain *curM* ST gene. The restriction enzyme digestion patterns of these cosmids were found to be similar, suggesting that they share the DNA fragments from the digested genomic DNA of *L. majuscula*. The two cosmids, pLM19 and pLM14, were further sequenced by primer walking in the CurM TE and the downstream regions. Our sequencing results confirmed that the previously published sequences (*3*) of CurM TE and also its downstream regions in pLM17 are not from the curacin A pathway, possibly due to a DNA rearrangement.

**Table 5-1. Primers for expression constructs.**

| Primer name | Sequences (5' to 3') |
|---|---|
| CurM-ACP-for | CATATGACAGACGAACGCATTTTAG (*Nde*I) |
| CurM-ACP-rev | GCGGCCGCTAAGCTTGTTGGAGATGG (*Not*I) |
| CurM-ST-for | CATATGATCTTTGCAACCAAAAGTTCA (*Nde*I) |
| CurM-ST-rev | GGATCCTTATTGAGGCTGTTGATTTGTCG (*BamH*I) |
| CurM-TE-for | CATATGCAAGTCTCTACAACTCCCT (*Nde*I) |
| CurM-TE-rev | GCGGCCGCGGATGTTAAGATAAGTGATGC (*Not*I) |

The restriction sites are underlined.

**Plasmid construction and site mutagenesis.** The primers used for the plasmid construction and site-directed mutagenesis were listed in table 5-1. CurM ACP, ST and TE genes were amplified from cosmid pLM14 or pLM19. The CurM ACP, ST-TE and ACP-ST-TE genes were inserted into pET24b plasmid with *Nde*I and *Not*I restriction sites. The CurM ST gene was inserted into pET28b plasmid with *Nde*I and *BamH*I restriction sites. All the constructs were verified by DNA sequencing.

**Protein overexpression.** *E. coli* BL21 (DE3) was transformed by the pET24b::CurM ACP, and pET24b::CurM TE plasmids to overexpress C-terminal His-tagged proteins, and by pET28b::CurM ST plasmid to overexpress N-terminal His-tagged protein. CurM ACP was overexpressed in the apo form. Cells were grown at 30$^{o}$C to an OD (590 nm) = 0.5-0.6, and then cooled to 15$^{o}$C prior to the addition of 1 mM isopropyl-β-D-galactopyranoside (IPTG). The cultures were grown at 15$^{o}$C for another 18-20 h before harvesting.

**Protein purification.** Protein purifications were performed at 4$^{o}$C. Generally, the first step Ni-affinity purifications for all His-tagged proteins were performed under the same conditions. *E. coli* cells were harvested by centrifugation (4,000 g, 15 min, 4$^{o}$C), resuspended in the ice cold lysis buffer A (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol) and disrupted by sonication on ice. The cell debris was removed by centrifugation at 15,000 g for 50 min. The supernatant was gently removed and loaded onto the 5 ml HisTrap column (GE Healthcare) preequilibrated with lysis buffer A. The resin was washed successively with ~ 10 column volumes of the washing buffer B (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol) to remove nonspecifically bound contaminants. Bound proteins were eluted with imidazole by a linear gradient of the elution buffer C (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 250 mM imidazole, 20% glycerol). The eluate was checked by SDS-PAGE for purity, pooled and concentrated using Amicon Ultra-15 (10 kDa or 5 kDa) centrifugal devices (Millipore). The concentrated eluate was loaded onto HiPrep 26/10 Desalting column equilibrated with the storage buffer D (50 mM PBS buffer, pH 7.5, 200 mM NaCl, 20% glycerol). The fractions were pooled, concentrated, flash-frozen in 50-200 μl aliquots in liquid $N_2$, and stored at -80$^{o}$C for further use. The purity of the proteins was analyzed by SDS-PAGE and the protein concentrations were determined using the Bradford assay

(Bio-Rad).

**The preparation of CurM ACP-linked model substrates.** The model substrates were firstly coupled to CoASH to generate CoA thioesters, which were loaded onto (apo) CurM ACP using sfp or svp protocol described previously (*9*).

**HPLC analysis of the ST-catalyzed sulfonation on the model substrates.** HPLC analysis of the ST-catalyzed sulfonation was performed using XBridge C18 column (4.6 x 250 mm, 5 μm, Waters) on the Gold HPLC system equipped with an autosampler and controlled by 32 Karat software (Beckman Coulter). The CoA samples were eluted with a linear gradient from 3% to 100% of MeOH/$H_2O$ (10 mM $CH_3CO_2NH_4$). The SNAC samples were eluted with linear gradient from 5% to 100% of $CH_3CN/H_2O$ (0.1% $CF_3COOH$). For the sulfonation assays, 200 μM CoA or SNAC model substrates were incubated with 20 μM CurM ST, and 1 mM PAPS in 50 mM Tris-HCl, pH 7.5, for 2 hours at room temperature. DMSO (1%) was added to increase solubility of SNAC substrates. The reaction mixtures were filtrated by Microcon YM-10 (Millipore), and analyzed by HPLC.

**HPLC analysis of the TE-catalyzed sulfonation on the model substrates.** The HPLC analysis conditions for the TE assays were similar to those of the ST assays described above. The CoA samples were eluted with a linear gradient from 3% to 100% of MeOH/$H_2O$ (10 mM $CH_3CO_2NH_4$). The SNAC samples were eluted with linear gradient from 5% to 100% of $CH_3CN/H_2O$ (0.1% $CF_3COOH$). For the TE assays, 200 μM CoA or SNAC model substrates were incubated with 10 μM CurM TE in 50 mM Tris-HCl, pH 7.5, for 1 hours at room temperature. DMSO (1%) was added to increase solubility of SNAC substrates. The reaction mixtures were filtrated by Microcon YM-10 (Millipore), and analyzed by HPLC.

**Analysis of CurM ACP samples by ESI-FTICR-MS**. CurM ACP samples were prepared by loading the reaction mixtures on Source 15PRC reverse phase column. The proteins were eluted with a linear gradient from 30% to 70% $CH_3CN$ (0.05% HCOOH and 0.05% $CF_3COOH$)/$H_2O$ (0.05% HCOOH and 0.05% $CF_3COOH$). Mass spectrometric analysis was performed with an actively shielded 7 Tesla

quadrupole-Fourier transform ion cyclotron resonance mass spectrometer (APEX-Q, Bruker Daltonics, Billerica, MA). Target analytes were diluted in an electrospray solution (1:1 $CH_3CN$:$H_2O$ with 0.1% HCOOH) and directly infused into an electrospray ionization (ESI) source (Apollo II, Bruker Daltonics) at a flow rate of 70 μL/h and a voltage of - 3.8 kV. A counterflow of hot (240°C) nitrogen gas was applied to assist desolvation of ESI droplets. For accurate mass determination, up to 10 picomoles of apomyoglobin (Sigma, St. Louis, MO) was spiked into the ESI solution as internal calibrant. Multiply protonated ions generated from ESI were externally accumulated in a hexapole for 1 s and transferred via high voltage ion optics to the ICR cell for analysis. All data were acquired with XMASS software (version 6.1, Bruker Daltonics) in broadband mode from m/z = 200 to 2000 with 512k data points and summed over 20-30 scans. Mass spectra were analyzed with the MIDAS analysis software (*10*). When needed, external frequency-to-m/z calibration was performed with a two-term calibration equation (*11*) using two calibration standards (m/z = 622.02895 and 922.00979, from the calibration mix G2421A, Agilent Technologies, Palo Alto, CA). For infrared multiphoton dissociation (IRMPD), precursor ions were mass-selectively accumulated in a hexapole with a 3-5 m/z quadrupole isolation window, transferred to the ICR cell, and irradiated for 100-300 ms by 10.6 μm photons at 10 W laser power (25 W $CO_2$ laser, Synrad, Mukilteo, WA) for 30-50 scans.

## 5.6 REFERENCES


1. C. T. Walsh, *Acc. Chem. Res.* **41**, 4 (2008).
2. J. V. Rossi, M. A. Roberts, H. D. Yoo, W. H. Gerwick, *J. Appl. Phycol.* **9**, 195 (1997).
3. Z. X. Chang *et al.*, *J. Nat. Prod.* **67**, 1356 (2004).
4. E. Chapman, M. D. Best, S. R. Hanson, C. H. Wong, *Angew. Chem. Intern. Ed.* **43**, 3526 (2004).
5. Y. Kakuta, L. G. Pedersen, C. W. Carter, M. Negishi, L. C. Pedersen, *Nat. Struct. Biol.* **4**, 904 (1997).
6. J. B. Bonanno *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12896 (2001).
7. S. S. Choi, Y. A. Hur, D. H. Sherman, E. S. Kim, *Microbiol.-SGM* **153**, 1095 (2007).
8. Z. X. Chang *et al.*, *Gene* **296**, 235 (2002).
9. L. C. Gu *et al.*, *J. Am. Chem. Soc.* **128**, 9014 (2006).

10. M. W. Senko, J. D. Canterbury, S. H. Guan, A. G. Marshall, *Rap. Comm. Mass Spectr.* **10**, 1839 (1996).
11. E. B. Ledford, D. L. Rempel, M. L. Gross, *Anal. Chem.* **56**, 2744 (1984).

# CHAPTER 6
# SUMMARY

## *6.1 NATURAL PRODUCT ASSEMBLY LINES IN EVOLUTION*

Natural product biosynthetic pathways are comprised of gene sets whose architecture reflects an ongoing evolutionary process. In bacterial type I PKS and NRPS systems, the metabolic pathway includes a series of modules that specify the nature of the starter unit, extender unit and termination processes to generate a core molecule. Within this biosynthetic framework numerous genetic determinants ultimately translate into the order, nature and number of catalytic domains within the PKS or NRPS multifunctional proteins. In turn, this determines the chemical outcome relating to each step during polyketide or non-ribosomal peptide assembly. As chain elongation progresses, keto group processing reactions determine the fate of the β-carbonyl group, stereochemistry of the resulting hydroxyl, or configuration of the double bond. In essence, each reaction operates independently, yet the catalytic event driving it has evolved within the scope of an extremely complex biosynthetic machine whose parts operate together efficiently.

The driving forces that determine co-evolution of multiple catalytic domains or enzymes within a secondary metabolic pathway remain largely unknown. Thus, although the biological role of many chemically diverse natural products is understood with respect to human cell targets, their natural or indigenous targets may never be definitively revealed. On the other hand, we assume that pathways of secondary metabolism evolve and expand by mechanisms similar to those recognized for primary metabolic pathways. However, tracing the ancestral forms of multiple genes as a functional collective is elusive, especially when they are dispersed in the genome. Biosynthetic genes from microbial hosts are usually clustered in their genomes, and as such are ideal for

evolutionary studies (*1*).

## 6.2 DIFFERENT TYPES OF MEATABOLIC PATHWAY EVOLUTION

The curacin A pathway enabled us to witness the innovation on a polyketide assembly line by different types of metabolic pathway evolution. This pathway is largely constructed by the canonical PKS and NRPS modules. However, in its chain initiation, extension and termination modules, some unusual catalytic elements were incorporated and catalyze the unprecedented reactions on the polyketide assembly line. These interesting catalytic enzymes and domains in canonical PKS module offer us a precious chance to reveal the evolution events that enabled the formation of curacin A pathway. The biochemical, bioinformatic and structural studies described in this dissertation demonstrated that these different types of pathway evolution are widely employed and significantly promote the chemical diversity of natural products.

### 6.2.1 Insertion of Gene Assembly by Homologous Recombination

The type I PKS and NRPS are highly modularized biosynthetic machinery. They serve as a paradigm for secondary metabolic systems that are successful in natural selection to expand chemical diversity. It is well know that the evolution of metabolic pathways proceeds by patterns of gene duplication, insertion and mutation. Recent bioinformatic studies suggest that some highly conserved sequences in PKS modules can facilitate domain replacement by homologous recombination (*2, 3*). As such, the conserved regions in these highly modularized systems give us a good reason to explain why these complex biosynthetic machines are chosen by nature to expand chemical diversity in secondary metabolism. In chapter 4, our comparative studies of the Cur and Jam pathways indicate that a whole gene assembly is inserted into polyketide pathway by the “di-AT” replacement. This type of the replacement might frequently happen in PKS pathway to achieve pathway expansion and contraction via an economic way. As more and more genome sequencing data become available, we will have enough samples to study the roles of homologous recombination in construction of new biosynthetic pathways. We also expect that this genetic strategy can be employed to the pathway

engineering in laboratory.

### 6.2.2 Recruitment of New Enzymes to Change Biochemical Schemes

In this dissertation, there are two perfect examples to show how new enzymes are recruited into the biosynthetic pathway and then significantly revise the biochemical schemes. In chapter 4, the Cur and Jam Hal domain are revealed to be introduced into the HMG enzyme assembly to catalyze the chlorination of (*S*)-HMG-$ACP_3$. This chlorination step revised the scheme of HMG β-branching modification. Without the chlorination step, an isovaleryl-$ACP_3$ is generated, but after the chlorination, the Cur ER catalyzes a cyclopropanation step instead of α,β saturation to form 2-methylcyclopropane-1-carboxyl-$ACP_3$. In chapter 5, another example is the terminal olefin formation in the chain termination of curacin A pathway. A sulfotransferase is incorporated into the chain termination module, and catalyzes the β-hydroxyl sulfonation to generate a good leaving group. Without the sulfotransferase, the thioesterase just cleaves the acyl chain from ACP to yield the product with a carboxylic end. Instead, the sulfonated product undergoes a decarboxylative chain termination to produce the terminal olefin. These two elegant biochemistries show us the tremendous capability of nature to innovate biosynthetic mechanisms from the existing ones.

### 6.2.3 Development of New Functions from Old Enzyme Scaffolds

The biosynthetic enzymes in secondary metabolism show significant substrate tolerance and catalytic promiscuity (*4, 5*), which can facilitate the generation of new functions from the same protein scaffold. In chapter 2, the $GNAT_L$ domain in the chain initiation module of curacin A pathway was proved to be a member of GNAT superfamily. GNATs normally catalyze the *N*-acetyltransfer from acetyl-CoA to the primary amine groups on small molecules or proteins. However, the $GNAT_L$ domain was demonstrated to catalyze the unprecedented *S*-acetyltransfer and decarboxylation reactions. Similarly, in chapter 4, the Cur ER domain shows the highest sequence identity to the Jam ER domain. Unexpectedly, the Cur ER catalyzes a cyclopropanation step of the chlorinated substrate, but the Jam ER domain only catalyzes the canonical saturation reaction. It is proposed that some mutations happen in the active site of Cur ER to

remove proton source from the α-carbanion in favor of the cyclopropanation via an intramolecular nucleophilic substitution. The $GNAT_L$ and ER domains well exemplify the intriguing protein evolution events that develop new functions from the same enzyme scaffolds. Our undergoing structural studies will let us know more details about the sequence variation that can yield these new activities.

### *6.2.4 Diversification of Regiochemical Control for Pathway Ramification*

In chapter 4, we show that the change of the decarboxylation regiochemical control can diversify pathway to produce the distinct functional groups in the final products. In HMG enzyme cassette, the $ECH_2$ domain normally generates decarboxylation products with a α,β C=C. However, the regiochemical control of Jam $ECH_2$ domain is differentiated to form the β,γ C=C product which can not be reduced by Jam ER. Some residues in a hypervariable region are suggested to contribute to the $ECH_2$ regioselectivity. Thus, our results suggest that the $ECH_2$ regiochemical control might be easily effected by the mutations happen in this hypervariable region, which might be a facile way to diversify biosynthetic pathways to form distinct functional groups.

### *6.2.5 Synergized Co-Evolution in Multienzyme System*

In natural product assembly lines, the changes on upstream enzymes will evidently effect downstream enzymes. If the downstream enzymes cannot effectively respond to upstream biochemical variations, the whole pathway will be nullified. In chapter 4, the introduction of a chlorination step renders the functional diversification of the Cur ER domain. The chlorination step can increase hydrophobicity of the final product. Given the unusually high hydrophobicity of curacin structure, which is successfully constructed to serve its host, the increase of the hydrophobicity might not be beneficial for the biological activity of final product. The Cur ER domain seems to respond to this change by eliminating the chlorine group to increase the hydrophobicity. Thus, this could be nice example for synergized co-evolution in multi-enzyme system.

## *6.3 REFERENCES*

1. M. A. Fischbach, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4601 (2008).

2. H. Jenke-Kodama, T. Borner, E. Dittmann, *PloS Comp. Biology* **2**, 1210 (2006).
3. C. P. Ridley, H. Y. Lee, C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4595 (2008).
4. C. Khosla, Y. Y. Tang, A. Y. Chen, N. A. Schnarr, D. E. Cane, *Ann. Rev.Bioch.* **76**, 11.1 (2007).
5. M. B. Austin, P. E. O'Maille, J. P. Noel, *Nat. Chem. Biol.* **4**, 217 (2008).