Production of Terminal Alkenes in Natural Product Biosynthesis: Structural Studies of Sulfotransferase and Thioesterase Didomains

by

Jennifer Gehret McCarthy

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in The University of Michigan
2012

Doctoral Committee

Professor Janet L Smith, Chair
Associate Professor Bruce Allan Palfey
Professor David H Sherman
Associate Professor Raymond C Trievel
Associate Professor Zhaohui Xu

Acknowledgements

First, I would like to thank my mentor, Janet Smith, for her guidance, support, and patience throughout my time in her lab. Her love of protein structure is contagious and inspiring. I would also like to thank my fellow lab members, past and present. The lab was a great place to work, laugh, and learn for the past five years.

This work would not have been possibly without the help of our many collaborators: Dr. Liangcai Gu and Eli Eisman in Dr. David Sherman's lab at the University of Michigan, Dr. Sarang Kulkarni in Dr. Peter Wipf's lab at the University of Pittsburgh, and Dr. Lena Gerwick and Dr. Bill Gerwick at Scripps Intuition of Oceanography. Thank you for your interest and great contributions to this project. Thank you to everyone at GM/CA beamline at the APS, Argonne National Lab, where I collected my crystallography data.

I would like to thank the other members of my thesis committee: Dr. Ray Trievel, Dr. Bruce Palfey, Dr. Zhaohui Xu, and Dr. Matt Young (previous member). The guidance at my committee meetings was instrumental in moving this project forward. Thank you also to the Program in Biomedical Sciences (PIBS) and the Department of Biological Chemistry (especially Beth Goodwin) for all of the logistical and administrative help over the years.

Finally, I would like to thank my friends and family, especially my husband, Marc, and my parents, Bob and Diane, for your love, prayers and encouragement!

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3-OST, heparan sulfate D-glucosaminyl 3-*O*-sulftransferse

A, adenylation

AA, acyl activating

ACP, acyl carrier protein

APS, advanced photon source

AR, adaptor region

AT, acyltransferase

C, condensation

CMT, *C*-methyltransferase

CoA, coenzyme A

Cy, condensation/cyclization

DH, dehydratase

DTT, dithiothreitol

E, epimerization

ECH, dehydrating enoyl-CoA hydratase

ER, enoylreductase

FAS, fatty acid synthase

FOM, average figure of merit

GCMS, Gas chromatography/mass spectroscopy

GHMP, galactokinas, homoserine kinase, mevalonate kinase, phosphomevalonate kinase

GNAT, GCN5-related *N*-acetyltransferase

Hal, halogenase

HCS, HMG-CoA synthase

HLD, haloalkane dehalogenase

HPLD, high pressure liquid chromatography

I, intensity

IPTG, isopropyl β-D-1-thiogalactopyranoside

KR, ketoreductase

KS, ketosynthase

LCMS, liquid chromatography/mass spectroscopy

MDD, mevalonate diphosphate decarboxylase

MME, monomethyl ether

NRPS, nonribosomal peptide synthetase

OLS, olefin synthase

OMT, *O*-methyltransferase

PAP, 3'-phosphoadenosine 5'phosphate

PAPS, 3'-phosphoadenosine 5'-phosphosulfate

PCP, peptidyl carrier protein

PDB, protein data bank

PCC, Pasteur culture collection

PEG, polyethylene glycol

PKS, polyketide synthase

RBS, ribosome binding site

RMSD, root mean squared deviation

SAD, single-wavelength anomalous diffraction

SeMet, selenomethionyl

SER, surface entropy reduction

ST, sulfotransferase

TB, terrific broth

TE, thioesterase

TLC, thin layer chromatography

THF, tetrahydrofuran

TSS, transcription start site

WT, wild type

Abstract

This thesis investigates a generalized biosynthetic scheme to produce hydrocarbons and natural products with a terminal alkene. Tandem sulfotransferase (ST) and thioesterase (TE) domains sulfonate a β-hydroxy group (ST) and hydrolyze, decarboxylate, and eliminate sulfate (TE) from an acyl carrier protein (ACP)-linked substrate to produce a terminal alkene instead of the carboxylic acid or macrolactone expected from a TE domain. Tandem ST-TE domains occur in the gene cluster for biosynthesis of the natural product curacin A (CurM) and in seven other bacterial genomes. Five of these convert free fatty acids to terminal-alkene hydrocarbons by an olefin synthase (OLS). To understand this novel biosynthetic strategy, the ST and TE domains from CurM (from *Moorea producens*) and OLS (from *Synechococcus sp.* PCC 7002) were investigated. Structural and biochemical investigations showed how the CurM and OLS ST and TE domains vary from canonical ST and TE folds to perform chemistry unique to the terminal-alkene producing systems. The CurM and OLS enzymes are similar to each other in structure and activity, but also exhibit subtle differences in substrate specificity, structure, and catalytic efficiency. The tandem ST-TE system is a promising chemical tool for the generation of new molecules or in the production of hydrocarbons for biofuel.

This thesis also includes structural and biochemical characterization of DmmA, an orphan haloalkane dehalogenase (HLD), originally annotated as CurN in the curacin pathway. DmmA has a larger active site than other studied HLDs and is a promising candidate for industrial applications for HLDs that act on larger substrates.

Chapter 1
Introduction

Background

Natural products are small-molecule secondary metabolites produced by bacteria, plants, and fungi. They exhibit an impressive array of chemical diversity and often potent bioactivity, which is frequently harnessed for therapeutics (1). Two classes of natural products, polyketides and non-ribosomal peptides, are synthesized in assembly-line-like biosynthetic pathways. Many unusual enzymes reside in natural product assembly-line pathways, and create the diverse collection of chemical functional groups found in natural products. The study of enzymes in natural product biosynthetic pathways can reveal new modes of catalysis, unique chemical transformations, and novel biosynthetic tools. One such reaction, the production of a terminal olefin, is the focus of this thesis.

Curacin A

The biosynthetic pathway producing the mixed polyketide/nonribosomal peptide curacin A contains the first example of the terminal-olefin producing chemistry (Figure 1.1) (2). Curacin A is produced by the marine filamentous cyanobacteria *Moorea producens* (3) (formerly *Lyngbya majuscula* (4)), which was originally isolated in Curaçao, Netherlands Antilles and is found in shallow tropical and subtropical marine environments around the world. *Moorea producens* grows in close association with other microorganisms (often residing on the cyanobacterial polysaccharide sheath (5)) and cannot be cultured in the absence of other microbes. Over 180 bioactive natural products including curacin A, barbamide (6), jamaicamide (7), scytonemin (8) and lyngbyatoxins (9) are attributed to *Moorea producens* (5). The recent genome sequence of *M. producens* 3L showed only eight natural product biosynthesis gene clusters, which

suggests that some natural products attributed to *M. producens* may be strain specific or produced by other microbes in the marine microbial consortium (5).

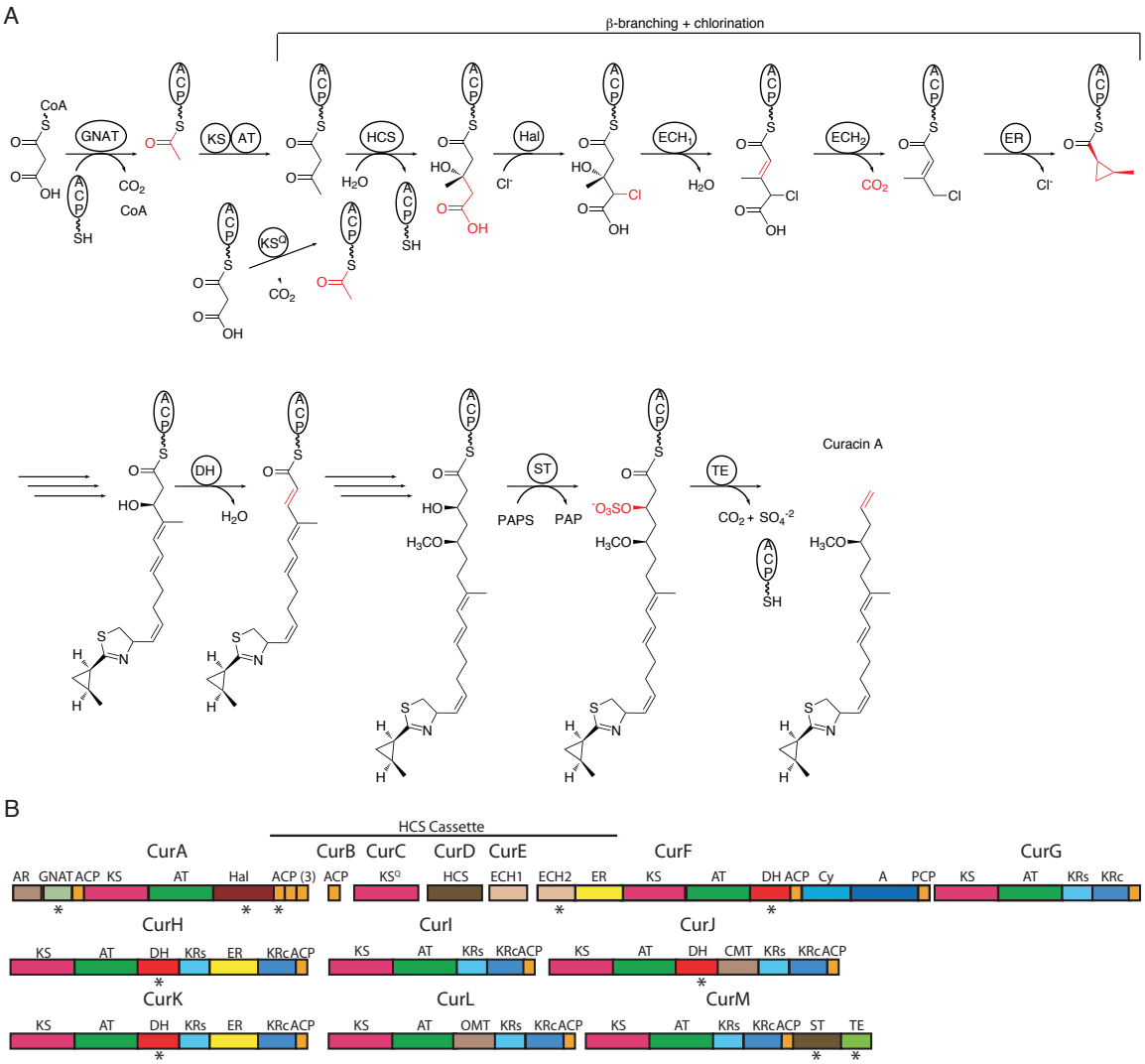


Figure 1.1 Curacin A biosynthesis

A. Curacin A biosynthesis with emphasis on reactions informed by protein structures. Changes to the polyketide intermediate are highlighted in red for each step. Abbreviations are: ACP, acyl carrier protein; GNAT, GCN5-related *N*-acetyltransferase; KS, ketosynthase; AT, acyltransferase; HCS, HMG-CoA synthase; Hal, halogenase; ECH_1, dehydrating enoyl-CoA hydratase; ECH_2, decarboxylating enoyl-CoA hydratase; ER, enoyl reductase; DH, dehydratase; ST, sulfotransferase; TE, thioesterase. **B.** Curacin A megasynthase. The annotated domains within the thirteen polypeptides (CurA – CurM) are colored by type. Asterisks indicate domains with published structures. Domains not listed in (A) are: AR, GNAT adaptor region; KR_C, ketoreductase catalytic domain, KR_S, ketoreductase structural domain; CMT, *C*-methyl transferase; OMT, *O*-methyl transferase; A, adenylation; Cy, condensation/cyclization; PCP, peptidyl carrier protein. Box sizes are proportional to domain sizes. Figure and caption from (10).

Curacin A is a potent antimitotic agent. It is a competitive inhibitor of colchicine-binding by tubulin and inhibits microtubule assembly (11). It has been shown to inhibit the growth of MCF-7 breast cancer cells (12). Even with knowledge of a potential therapeutic application, the natural function of curacin A is still unknown. Polyketides and nonribosomal peptides are known to be used by the producing bacteria against competitors within the microbial environment, as chemical signals between microbes (13), or as virulence factors in pathogens (14).

Biosynthesis of natural products

The curacin A biosynthetic pathway is a hybrid polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) assembly line (Figure 1.1B) (2). PKS and NRPS pathways are composed of modules, which extend and modify the growing intermediate in a sequential, ordered fashion. The genes for a single pathway are typically clustered in the genome and are frequently found in the same order that the gene products act in the assembly line (15). Each PKS module uses the ketosynthase (KS) and acyltransferase (AT) domains to extend the intermediate chain by two carbons from malonyl-CoA or methylmalonyl-CoA. The optional modifying domains can then reduce the β-carbonyl to a β-hydroxy (ketoreductase (KR)), dehydrate the β-hydroxy to an α,β-enoyl (dehydratase (DH)), and reduce the enoyl double bond to produce a saturated acyl chain (enoylreductase (ER)) (16, 17). Occasionally, other domains that make additional modifications, such as a C- or O-methyltransferase, are also present in the modifying section of the module. The presence or absence of modifying domains in each module allows different chemical groups to be produced in each step of the synthesis. PKS modules are similar to the mammalian fatty acid synthase (FAS) both in sequence identity of the individual domains and the organization of the domains within the entire module (18). NRPS modules use an adenylation doamain (A) to select and activate an amino acid and a condensation domain (C) to form the peptide bond that extends the chain by one amino acid (16). NRPS modules can create even greater chemical diversity by the inclusion of a methyltransferase, epimerization domain (E), or condensation/cyclization domain (Cy) (16).

PKS and NRPS modules can combine to create hybrid pathways in part due to a common intermediate-tethering strategy. Following the catalytic domains, each NRPS or PKS module contains a 10 kDa domain termed acyl carrier protein (ACP) in PKS or peptidyl carrier protein (PCP) in NRPS that covalently tethers the intermediate (16). A phosphopantetheinyl "arm" from Coenzyme A (CoA) is attached by a phosphopanetheinyltransferase to a conserved serine in the ACP or PCP. The substrate is tethered to the phosphopantetheine arm through a thioester linkage. The growing intermediate is passed between the carrier proteins (ACP or PCP) in each module until it is offloaded as a free small molecule at the pathway termination. Offloading typically occurs by a thioesterase (TE) domain, which follows the ACP or PCP in the final module. Typically, a TE domain catalyzes thioester hydrolysis to produce a carboxylic acid or a cyclic marcrolactone or macrolactam (Figure 1.2) (16).

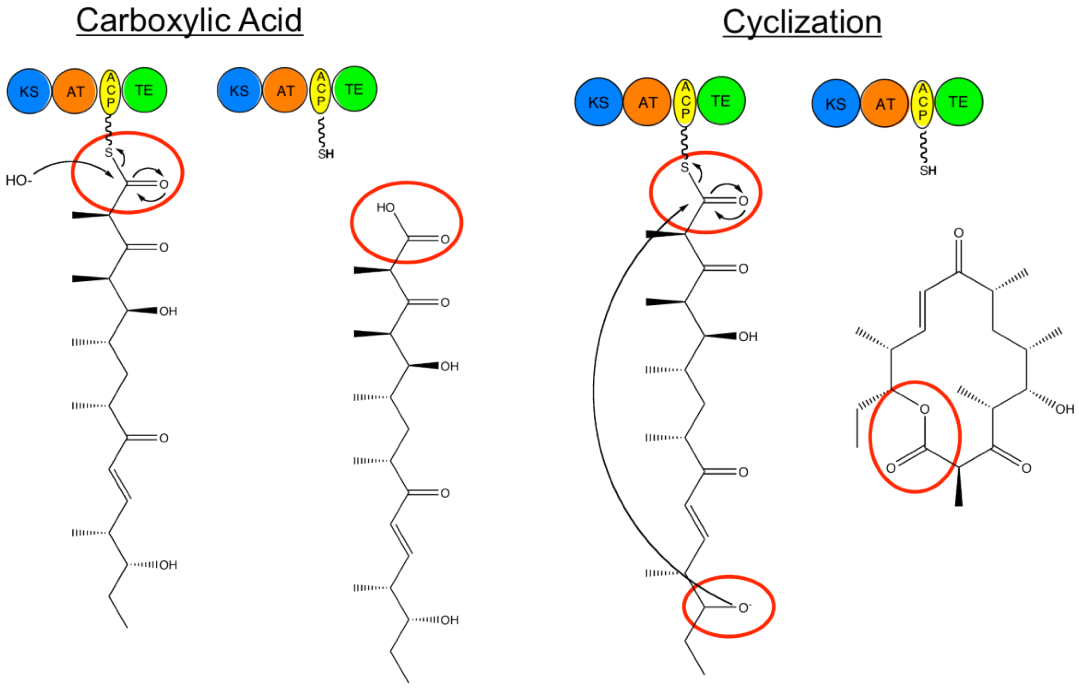


Figure 1.2 Canonical offloading using a thioesterase
In PKS pathways, a thioesterase (TE) domain is present at the C-terminus of the final module in the pathway. The TE offloads the free small molecule as a linear product with a carboxylic acid functional group or as a cyclic marcrolactone.

Biosynthesis of curacin A

Curacin A has several chemical moieties that are rarely seen in natural products including cyclopropane and thiazoline rings, an internal *cis* double bond and a terminal alkene. With the discovery of the curacin biosynthetic gene cluster (2), it was apparent that sequence analysis alone could not fully rationalize how curacin A was synthesized from the gene cluster. The curacin PKS/NRPS pathway contains seven standard PKS modules and one NPRS module, consistent with its polyketide backbone and thiazoline moiety (Figure 1.1) (2). Structural and biochemical studies of the non-standard domains in the pathway have elucidated the unusual biosynthetic strategies employed in curacin A biosynthesis. The biosynthesis is initiated by a rare GNAT (GCN5-related *N*-acetyltransferases), which catalyzes both malonyl-CoA decarboxylation and acyltransfer from CoA to ACP (19). This intermediate then enters the HSC cassette (CurA ACP_3 through CurF ECH_2-ER), which is occasionally employed in other PKS pathways to create a β-branch. The curacin pathway instead uses the HSC cassette to produce the terminal cyclopropyl moiety by the addition of an Fe^{2+}/O_2/α-ketoglutarate-dependent halogenase (CurA Hal), and a specialized ER within CurF (20-25) (Figure 1.1). Studies of the DH domains in the curacin pathway show how they can accommodate the bulky and stiff intermediates (26) and led to a hypothesis explaining the synthesis of the internal *cis* double bond. Additionally, the final module, CurM, catalyzes the offloading of the final product containing a terminal alkene (Figure 1.3A).

Terminal double bond offloading

The final PKS module in the curacin A pathway (CurM) contains a sulfotransferase (ST) and thioesterase (TE) domain C-terminal of the ACP domain. The module directly upstream of CurM, CurL, produces an intermediate with a 13-carbon acyl chain following the thiazoline. The N-terminal domains of CurM (KS-AT-KR-ACP) extend the acyl chain by two carbons (making a 15-carbon chain) and reduce the β-carbonyl to a β-hydroxy. The product curacin A has only 14 carbons following the thiazoline. Therefore, CurM ST and TE should catalyze a decarboxylation in addition to thioester hydrolysis and terminal double bond formation to generate curacin A (Figure 1.3A). CurM ST and CurM TE have low sequence identity to other studied STs and TEs and from sequence analysis alone it is unclear how they catalyze these reactions.

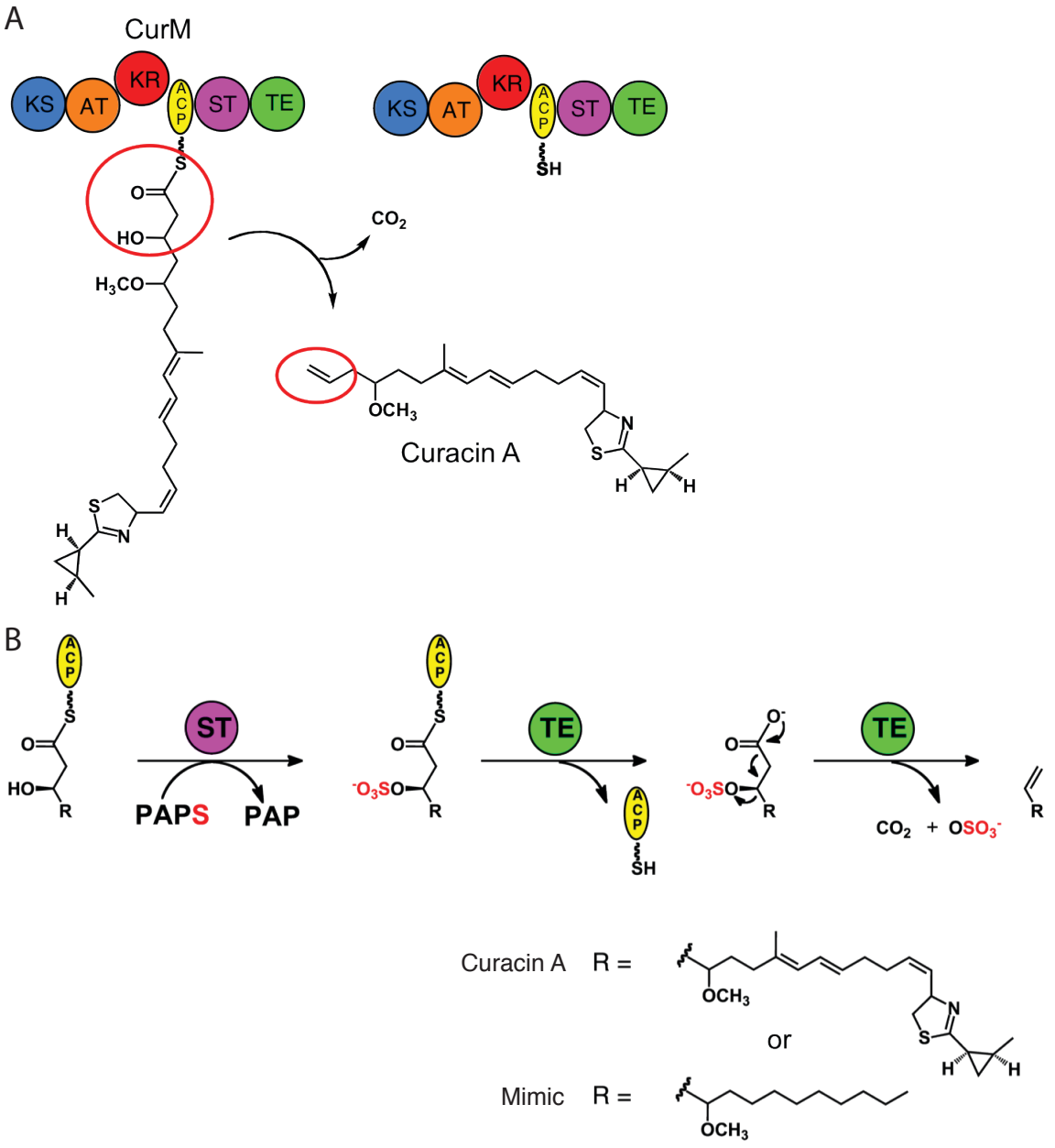


Figure 1.3 Offloading in curacin A biosynthetic pathway
A. CurM, the final module in the curacin A PKS/NRPS pathway. CurM contains sulfotransferase (ST) and thioesterase (TE) domains C-terminal of the canonical PKS domains, ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), and acyl carrier protein (ACP). CurM ST-TE decarboxylates and offloads the covalently tethered intermediate to produce curacin A. **B.** Mechanism of terminal double bond formation. CurM ST sulfonates the β-hydroxy substrate from the sulfonate donor PAPS (3′-phosphoadenosine 5′-phosphosulfate). CurM TE then hydrolyzes the thioester and decarboxylates and eliminates sulfate to produce a terminal alkene. Biochemical characterization was performed using a mimic of the natural substrate.

Biochemical studies were performed with CurM ST and CurM TE domains excised from the larger CurM module (Figure 1.3B) (27). The native substrate was expected to contain a β-hydroxy group generated by the N-terminal domains in the CurM module (KS-AT-KR). Chemically synthesized substrate mimics containing the β-hydroxy and C5-methoxy groups from the natural substrate were loaded on the excised CurM ACP. LCMS and GCMS were used to detect the ACP-tethered intermediates and the free small molecule products.

CurM ST was found to act first on the β-hydroxy group of the ACP-linked intermediate (27). CurM ST transfers a sulfonate using the donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS). CurM TE then acts upon the activated intermediate containing a β-sulfate to hydrolyze the thioester, decarboxylate, and eliminate sulfate, to yield the product olefin (Figure 1.3B). CurM TE shows great specificity for substrates with a β-sulfate, being 800x slower on the analogous β-hydroxy substrates. From the reaction of CurM TE with the β-sulfate substrates, two small-molecule products were detected, one containing a terminal double bond (similar to the curacin A) and one with the β-sulfate and a carboxylic acid (an off-pathway intermediate). CurM TE displayed decarboxylation and sulfate elimination activity in addition to the thioester hydrolysis typical of offloading TEs. The scheme in which the ST sulfonates the β-hydroxy to become a better leaving group is the first observation in biology of leaving-group activation by sulfonation.

Canonical sulfotransferases and thioesterases

Offloading in the curacin A pathway using CurM ST and CurM TE contrasts greatly with standard offloading in PKS or NRPS pathways. Typically, a TE domain catalyzes thioester hydrolysis by activation of a water to produce a carboxylic acid on a linear acyl or peptidyl chain or activates an intramolecular nucleophile to produce a marcrolactone, macrolactam, or cyclic peptide (Figure 1.2) (16). The offloading TEs, like the rest of the TE family, have an α/β hydrolase fold and use a nucleophile-His-Asp/Glu catalytic triad to catalyze the hydrolytic chemistry. The α/β hydrolase core is conserved across all of the offloading TEs, but the lid structure is variable. The lids of PKS offloading TEs are permanently closed by dimer contracts, which create a tunnel with the active site at the center (28-32). In contrast, the lids of the monomeric NRPS

offloading TEs are flexible (33-38). Non-offloading (Type II) TEs, which have an editing or specialized function, are also monomeric with flexible lids (39, 40).

Sulfotransferases are rarely observed in PKS or NRPS pathways. In cases where they are found, they appear to produce sulfated natural products (41-43). STs catalyze sulfonate transfer from the donor PAPS to a hydroxy or amine acceptor. Catalysis occurs using a general base to deprotonate the hydroxy or amine, which attacks the PAPS sulfonate group. STs have been most extensively studied in mammals where they are classified by their cellular compartmentalization (44). Cytoplasmic STs act upon small molecules such as neurotransmitters, hormones, and drugs (44). Golgi-resident STs act upon macromolecules (carbohydrates and proteins) that are secreted from the cell (45, 46). In bacteria, STs are also involved in sulfolipid biosynthesis (47). Despite the varied known roles for sulfonation, sulfonation for functional group activation, such as in the curacin A pathway, has not been previously observed.

Other instances of ACP-ST-TE tridomains

Genes encoding the ACP-ST-TE tridomains were found in seven other bacterial genomes (Figure 1.4A) (10). Two genes encoding ACP-ST-TE occur in the context of a PKS or NRPS gene cluster. One cluster is from the marine myxobacteria *Haliangium ochraceum* (48), which produces the haliangicins (49, 50), natural products containing a terminal double bond. However, the predicted product from the gene cluster containing the ACP-ST-TE sequence is unrelated to the haliangicins, which are likely produced by a separate pathway (Figure 1.5F). The second gene cluster that contains the ACP-ST-TE sequence is from the entomopathogenic proteobacteria *Pseudomonas entomophila* (Figure 1.5E) (51). In both *H. ochraceum* and *P. entomophila*, the natural product predicted to be produced from the pathways has not been detected in the microbe. It is expected, however, that the pathways produce natural products containing a terminal double bond.

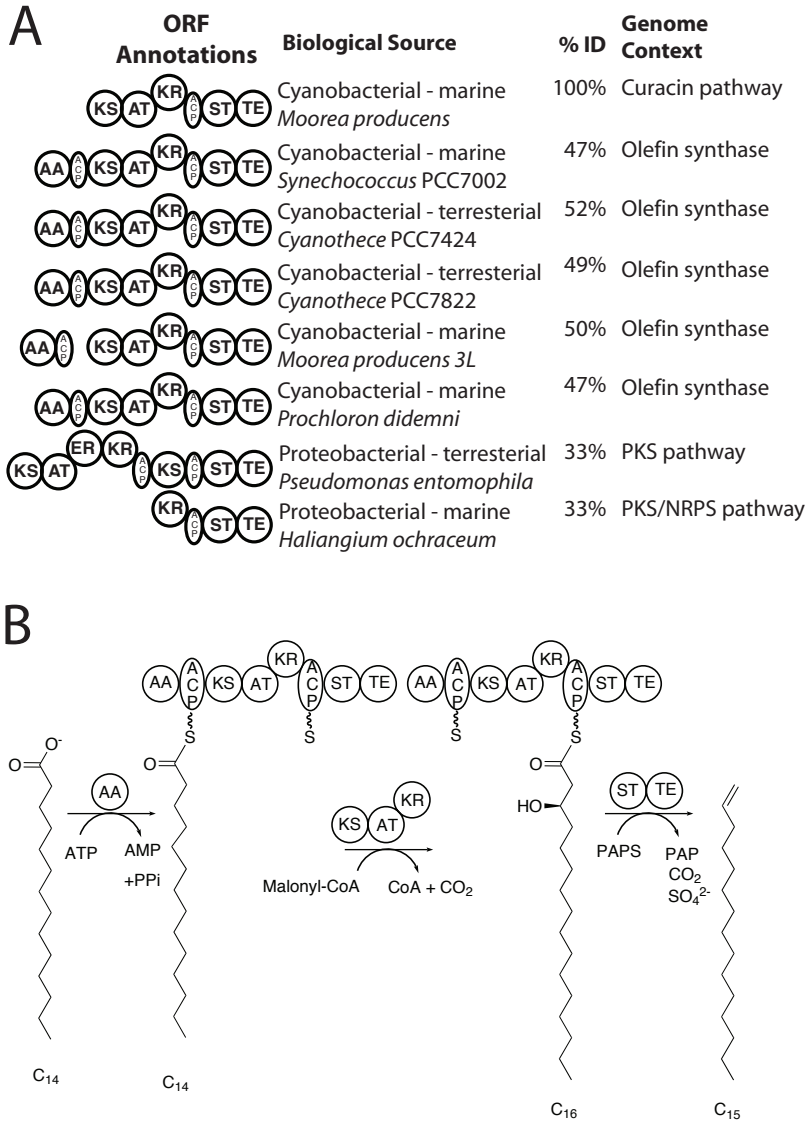
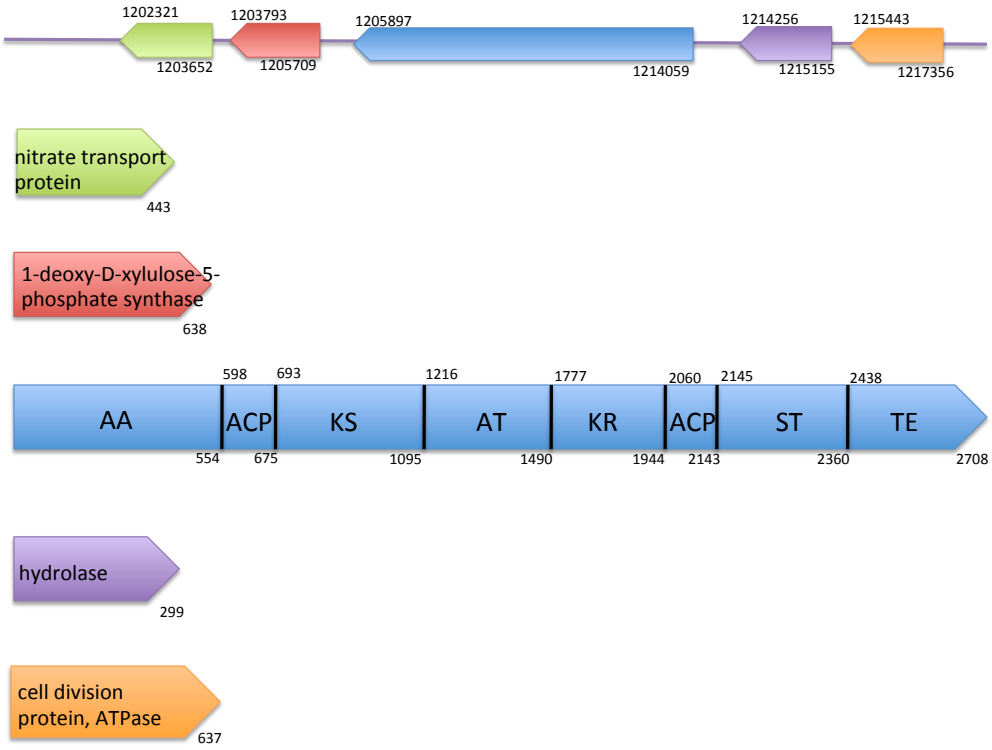


Figure 1.4 Distribution of CurM ST-TE homologues

A. ORFs encoding ACP-ST-TE domains. The predicted domains are depicted for all of the sequences encoding sequential ACP-ST-TE domains. GenBank entries are: *Moorea producens* (ACV42478), *Synechococcus* PCC 7002 (YP_001734428), *Cyanothece* PCC 7424 (YP_002377174), *Cyanothece* PCC 7822 (ZP_03153601), *Moorea producens* 3L (ZP_08425908), *Prochloron didemni* (AEH57210), *Pseudomonas entomophila* L48 (YP_610919), *Haliangium ochraceum* DSM 14365 (YP_003265308). **B.** Predicted pathway to hydrocarbons based on sequence annotation. The acyl-activating (AA) domain is predicted to load a fatty acid onto the adjacent ACP, the KS-AT-KR to extend and reduce the acyl-ACP to produce a β-hydroxy substrate for the ST-TE. Figure and caption from (10).

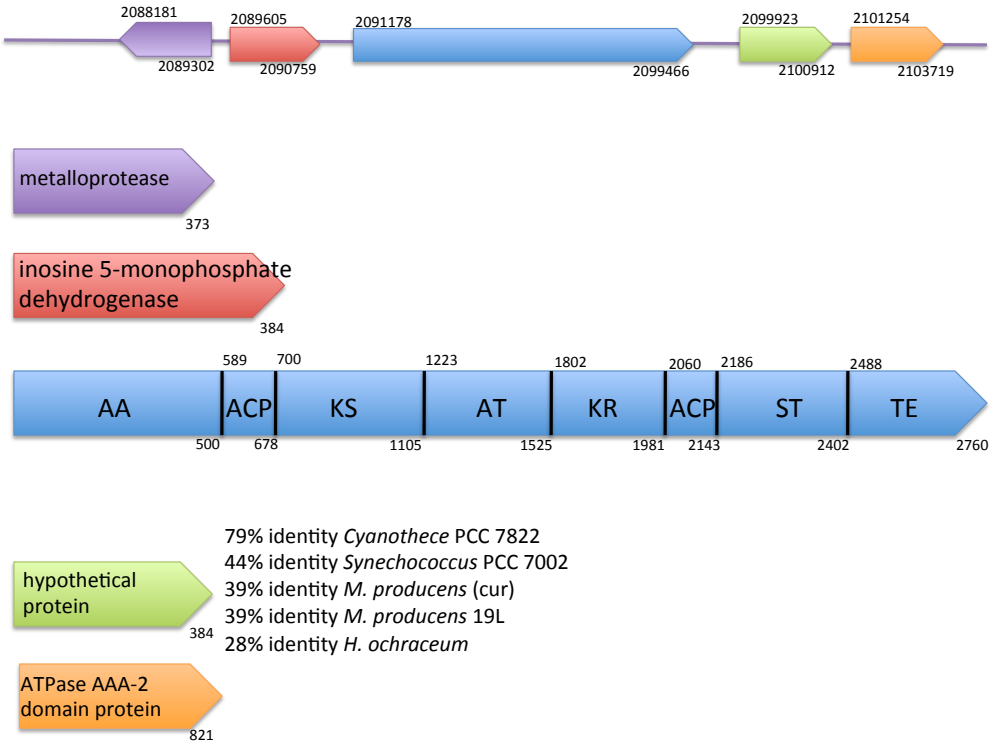
A *Synechococcus* PCC 7002

47%

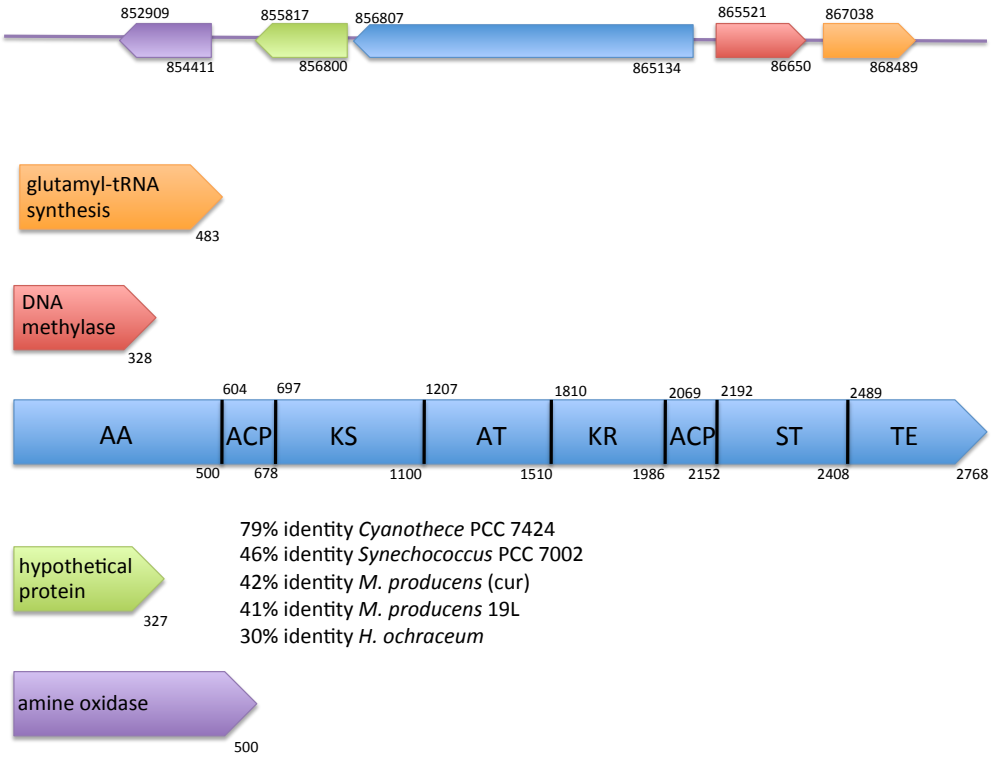


B *Cyanothece* PCC 7424

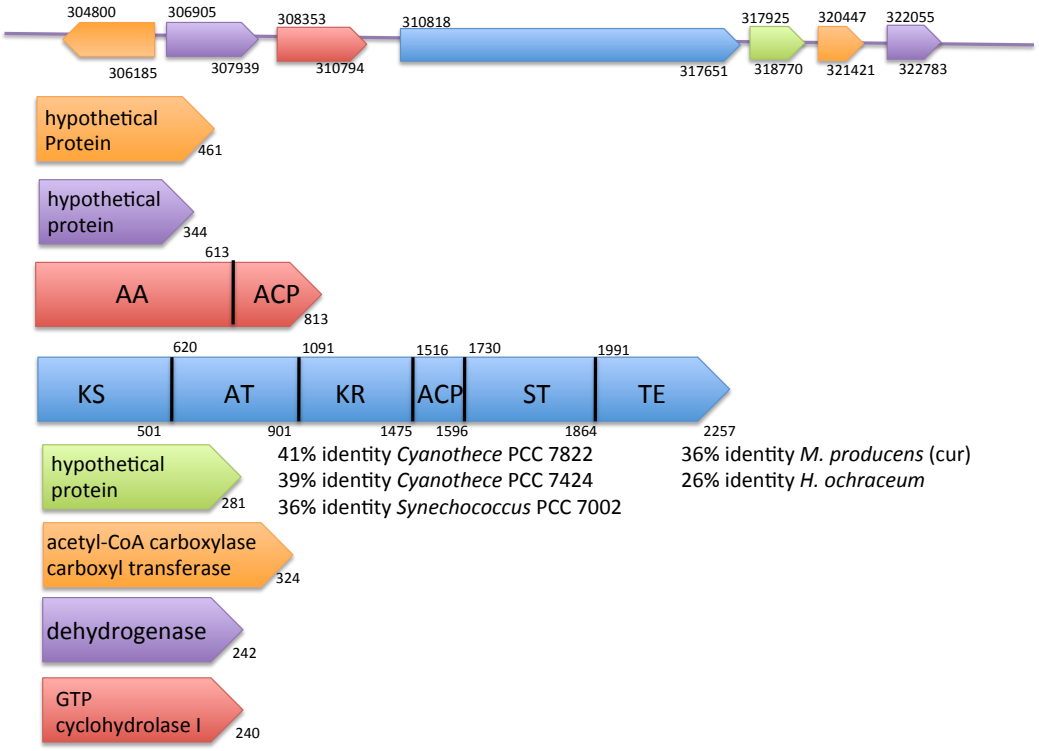
52%



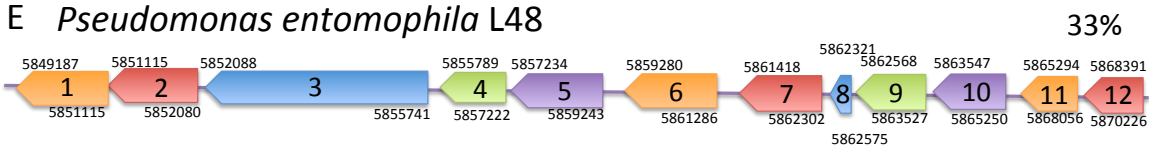
C *Cyanothece* PCC 7822 49%

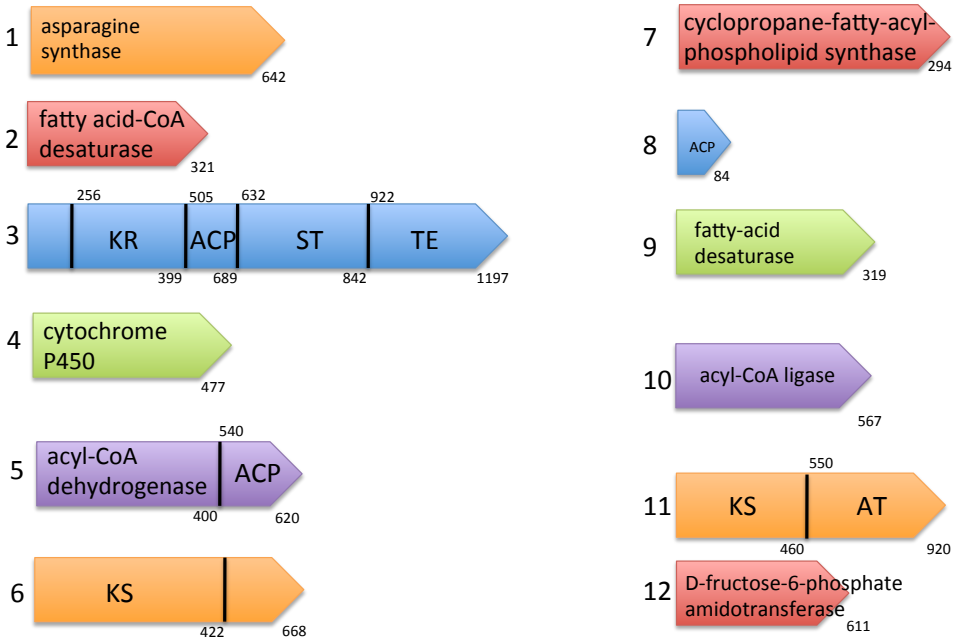


D *Moorea producens* 3L 50%

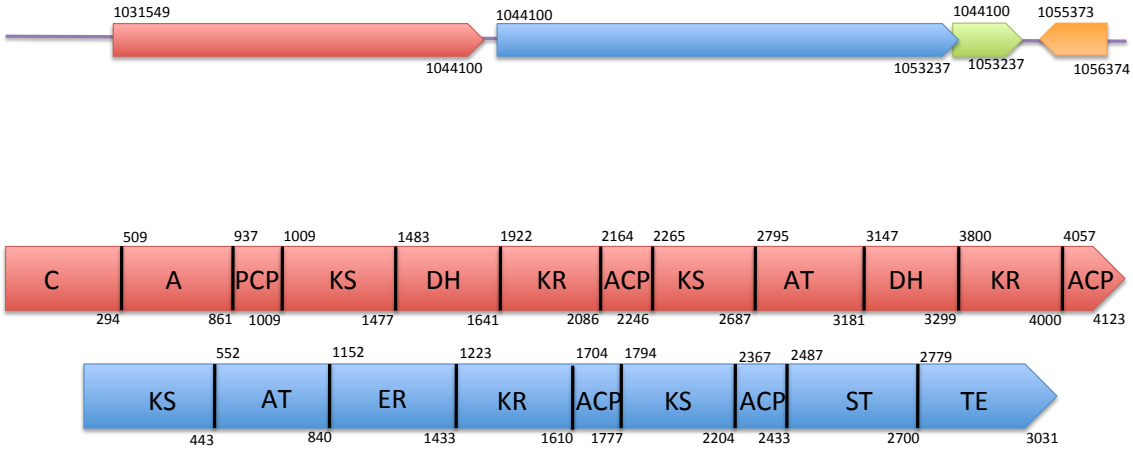


E *Pseudomonas entomophila* L48





F *Haliangium ochraceum* DSM 14365



hypothetical protein

putative transposase
333

30% identity *Cyanothece* PCC 7822
28% identity *Cyanothece* PCC 7424
26% identity *M. producens* L19
26% identity *M. producens* (cur)
26% identity *Synechococcus* PCC 7002

Figure 1.5 Genomic context for genes encoding ACP-ST-TE
Percent identity to CurM ACP-ST-TE is listed in the top right corner. The genomic context is illustrated on top, numbers represent the location of the gene within the genomic contig. Annotation (from NCBI) of the corresponding gene products are depicted below with numbers indicating amino acids. *P. didemni* is not shown because the genomic region surrounding the *ols* gene is not available in the database.

The remaining five ACP-ST-TE sequences are not in the context of natural product gene clusters but are stand-alone systems, all with similar domain arrangements (Figure 1.5A-D) (10). The five gene products have virtually identical domain arrangements and are termed olefin synthases (OLS) (52) instead of polyketide synthases (PKS). In every case, the coding sequences for the acyl activating (AA) and ACP domains are present upstream of the sequence coding for the domains found in CurM (KS-AT-KR-ACP-ST-TE). The AA domain resembles those shown to ligate a free fatty acid to ACP through an acyl-AMP intermediate (53). The KS-AT-KR can extend the activated fatty acid by 2 carbons and reduce the β-carbonyl to a β-hydroxy. The intermediate can then proceed through the decarboxylative offloading process catalyzed by the ST-TE to result in a terminal alkene hydrocarbon, one carbon longer than the starting fatty acid (Figure 1.4B). The OLS system is distinctive in converting a fatty acid with an even number of carbons (n) to a hydrocarbon with an odd number of carbons (n+1).

In all five instances, the olefin synthase gene (*ols*) was found in cyanobacterial genomes. One of the cyanobacteria, *Prochloron didemni*, lives symbiotically with tropical ascidians (marine invertebrate filter feeders). The *ols* gene was found in *P. didemni* samples associated with different ascidian species collected from sites spanning 5500 km of the tropical Pacific (54). Another *ols* gene is found in *Moorea producens* (5), the same species that produces curacin A. Two *Cyanothece* (PCC 7424 and PCC 7822) (Figure 1.5B-C) and one *Synechococcus* (PCC 7002) (Figure 1.5A) species also contain the *ols* gene. The olefin synthase activity of the *ols* gene-product in *Synechococcus sp*. PCC 7002 was confirmed through gene-deletion and feeding studies (52).

Chemistry related to ST-TE olefin production

Terminal-alkene hydrocarbons with an odd number of carbons appear in other microbes that lack a sequenced genome. It is possible that an ST-TE system also

produces hydrocarbons in these species. One example is the microalga *Botyrococcus barunii* (55), which produces terminal unsaturated dienes and trienes with an odd number of carbons (27, 29, or 31). Feeding experiments have shown that hydrocarbon production starts with C_{14} or C_{16} fatty acids, proceeds through many 2-carbon elongation steps, and terminates with a β-activated decarboxylative process similar to the ST-TE (56). Other alga such as *Isochrysis galbana* and *Emiliania huxleyi* produce similar terminal unsaturated long chain hydrocarbons (57), possibly through the same mechanism.

Terminal alkenes in isoprenoids are synthesized by a related catalytic strategy using phosphate in place of sulfate. Mevalonate diphosphate decarboxylase (MDD) phosphorylates a hydroxy group in mevalonate-5-diphosphate. MDD then decarboxylates and eliminates phosphate to produce isopentenyl diphosphate, containing a terminal double bond (58, 59). Structurally, MDD belongs to the GHMP superfamily (galactokinase, homoserine kinase, mevalonate kinase, phosphomevalonate kinase) of small-molecule kinases (60). Although the chemistry is similar to ST-TE catalysis, MDD does not catalyzed thioester hydrolysis and does not resemble the ST-TE proteins.

ST-TE olefin offloading as a chemical tool

The ST-TE offloading strategy generates hydrocarbons when coupled with the AA fatty acid loading domain in the olefin synthase (Figure 1.4B). The hydrocarbons produced by OLS are appealing biofuels compatible with existing engines, fuel storage, and distribution systems (61, 62). Terminal alkene hydrocarbons make ideal diesel fuel (C_9-C_{23}) or jet fuel (C_8-C_{16}) (63, 62). Consequently, OLS has the ability to be developed for biofuel production, providing an attractive alternative to petroleum-derived fuel, potentially addressing energy-security and environmental concerns (64). OLS could be used in an endogenous bacterium or engineered into a recombinant host. *E coli*, for example, has already been engineered to overproduce the fatty acid substrates for OLS (65). The cyanobacteria with endogenous *ols* genes are also promising since they are able to capture solar energy through photosynthesis.

The ST-TE enzymes also present an opportunity for natural product structural diversification through combinatorial biosynthesis. Genetic manipulation of PKS or NRPS pathways allows the production of bioactive compounds with novel chemical functionalities that are inaccessible or challenging to obtain by chemical synthesis (66,

67). The ST-TE should be able to replace a normal offloading TE in any pathway that produces a β-hydroxy in the final module. The capability to create an alkene functional group during the offloading process instead of the carboxylic acid or macrolactone will lead to an increase in the diversity of small molecules available to be screened as therapeutic targets, reaching rates much greater than what is currently available in chemical synthesis.

The lost haloalkane dehalogenase CurN/DmmA

When the curacin A biosynthetic gene cluster was initially characterized, an additional open reading frame, *curN*, encoding a haloalkane dehalogenase (HLD) was found following *curM* at the end of the gene cluster (2). DNA encoding 64 amino acids at the CurN N-terminus overlapped with DNA encoding the CurM C-terminus in a different reading frame. Later resequencing of the cluster showed that *curM* had a different 3′ end and *curN* was not in the cluster (27). The *Moorea producens* genome also does not contain *curN* (5), which was likely produced by another organism in the microbial consortium. CurN was then renamed DmmA (dehalogenase A from a marine microbe) since it is known not to belong to the curacin pathway or *M. producens* (68).

Haloalkane dehalogenases (HLD) remove halogens from alkanes by hydrolysis, producing an alcohol, a halide ion, and a proton. HLDs have been developed in industry for over 20 years for applications such as degradation of environmental pollutants (69, 70), biocatalysis (71), remediation of chemical weapons (72), biosensing (73, 74), and cellular imaging (75). Even though the source organism of DmmA is uncertain it is an appealing target for chemical tool development.

Thesis overview

The following chapters present the biochemical and structural characterization of the ST and TE domains from a PKS (curacin) and OLS (*Synechococcus sp*. PCC 7002) pathway. Both pathways present novel variations on the ST and TE folds previously characterized. Structural analysis coupled with assays of protein variants and chimeras shows how the canonical ST and TE folds are adapted to perform the unique functions of sulfonation, hydrolysis, and decarboxylative elimination. Investigation of the ST-TE didomain shows no direct channeling of the substrate between the two domains. This

thesis also includes an investigation of DmmA, which has haloalkane dehalogenase activity and a larger active site binding pocket than other HLDs.

Chapter 2

Terminal Alkene Formation by the Thioesterase of Curacin A Biosynthesis: Structure of a Decarboxylating Thioesterase

This chapter was published Gehret JJ, Gu L, Gerwick WH, Wipf P, Sherman DH, Smith JL (2011) Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. J Biol Chem 286:14445-14454.

Summary

Curacin A is a polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) derived natural product with potent anticancer properties generated by the marine cyanobacterium *Lyngbya majuscula*. Type I modular PKS assembly lines typically employ a thioesterase (TE) domain to offload carboxylic acid or macrolactone products from an adjacent acyl carrier protein (ACP) domain. In a striking departure from this scheme the curacin A PKS employs tandem sulfotransferase (ST) and TE domains to form a terminal alkene moiety. ST sulfonation of β-hydroxy-acyl-ACP is followed by TE hydrolysis, decarboxylation and sulfate elimination (22). With low sequence identity to other PKS TEs (<15%), the curacin TE represents a new thioesterase sub-family. The 1.7-Å curacin TE crystal structure reveals how the familiar α/β hydrolase architecture is adapted to specificity for β-sulfated substrates. A Ser-His-Glu catalytic triad is centered in an open active-site cleft between the core domain and a lid sub-domain. Unlike TEs from other PKSs, the lid is fixed in an open conformation on one side by dimer contacts of a protruding helix and on the other side by an arginine anchor from the lid into the core. Adjacent to the catalytic triad, another arginine residue is positioned to recognize the substrate β-sulfate group. The essential features of the curacin TE are conserved in sequences of five other putative bacterial ACP-ST-TE

tridomains. Formation of a sulfate leaving group as a biosynthetic strategy to facilitate acyl chain decarboxylation is of potential value as a route to hydrocarbon biofuels.

Introduction

Natural products display a remarkable chemical diversity, providing advantages for the producing plants and microbes to survive and thrive in particular ecological niches. These secondary metabolites and their derivatives have important applications as pharmaceuticals (76), and some have potential to be developed as biofuels (61). Gene clusters encoding assembly-line biosynthetic pathways for polyketide and polypeptide natural products are ubiquitous in bacterial and fungal genomes. Polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) pathways have a common modular organization in which intermediates tethered to carrier domains by a thioester linkage pass sequentially through modules of the assembly line. The final step in the assembly line is typically a thioesterase-catalyzed offloading from the final carrier domain to produce a carboxylate, macrolactone, or cyclic peptide.

In a notable exception to this offloading paradigm, the curacin A final product contains a terminal alkene moiety. Curacin A, from the marine cyanobacterium *Moorea producens* (formerly *Lyngbya majuscula* (4)), is a mixed polyketide/non-ribosomal peptide with antimitotic properties (12). The hybrid PKS/NRPS assembly-line pathway for curacin A (2) generates several unusual chemical groups in addition to the terminal alkene, including a cyclopropyl ring, a thiazoline ring, and a *cis* double bond. We have investigated the biosynthetic steps leading to several of these segments (21, 19, 22, 27, 26). Herein we investigate the structural basis for the unique offloading strategy leading to the terminal alkene in the curacin A molecule.

The curacin PKS has an unusual terminal module, the CurM protein, with a C-terminal tridomain comprised of an acyl carrier protein (ACP), a sulfotransferase (ST) and a thioesterase (TE) (Figure 2.1A). Annotations of both ST and TE were based on weak sequence similarity to characterized enzymes. The prediction of an ST within a PKS was unprecedented (2). STs are widely distributed and are known to have detoxification, hormone regulation or signaling functions (44). They catalyze transfer of a sulfonate group from the donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to a hydroxyl or amine of an acceptor small molecule or protein. CurM TE, although

identifiable as a thioesterase, does not resemble any of the previously established PKS or NRPS TE sub-families (77).

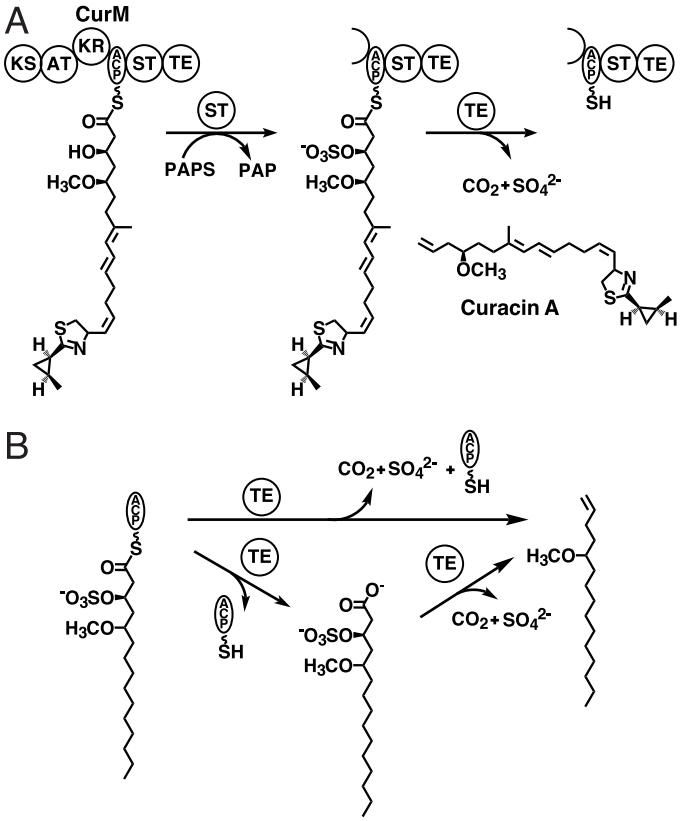


Figure 2.1 Chain termination in curacin A biosynthesis
A. Offloading reactions in the final module, CurM. Following extension by the CurM ketosynthase (KS) and acyltransferase (AT) and reduction by the ketoreductase (KR), ST catalyzes transfer of sulfonate to the β-hydroxyl followed by TE hydrolysis of the thioester with concomitant decarboxylation and sulfate elimination. The sulfonate group donor is the PAPS (3'-phosphoadenosine 5'-phosphosulfate) cofactor, which is converted to PAP (3'-phosphoadenosine 5'-phosphate). **B.** Experimental scheme for assay of CurM TE. Recombinant CurM ACP loaded with a substrate analog and sulfonated by ST was reacted with excised CurM TE. Activity was monitored by HPLC of the reaction mixture and detection of the holo-ACP product.

Using a simplified analog of the penultimate pathway intermediate, we recently demonstrated that offloading and terminal alkene formation require ST-mediated sulfonation of the β-hydroxyl group from the PAPS cofactor (27) (Figure 2.1A). This was the first observation of biological substrate activation by formation of a sulfate leaving group. CurM TE acts upon the β-sulfate intermediate to yield a decarboxylated product with a terminal double bond resulting from sulfate elimination (Figure 2.1B).

CurM TE catalyzes thioester hydrolysis 800-fold more slowly on the corresponding substrate bearing a β-hydroxyl group. The unprecedented requirement of a β-sulfate for thioester hydrolysis as well as the decarboxylation and sulfate elimination suggests a unique catalytic strategy and active site structure for the TE. Moreover, this curacin pathway decarboxylation strategy provides an opportunity to investigate a new biological route to hydrocarbon production from fatty acids. Thus, in addition to interest in curacin biosynthesis as a route to a potent anti-cancer compound (12), the pathway also has relevance to biofuel production.

Offloading TEs from many PKSs have been studied including the TEs of the pikromycin (Pik TE) (28, 30, 31), erythromycin (28), and tautomycetin (32) PKSs. PKS offloading TEs typically catalyze either hydrolysis to produce a linear carboxylic acid or the attack of an intramolecular hydroxyl to produce a large-ring macrolactone. PKS TEs are members of the α/β hydrolase superfamily with a catalytic-triad active site located at the top of an α/β core and covered by an α-helical lid sub-domain. All PKS offloading TEs of known structure are dimers in which two N-terminal α-helices in the lid form a lid-to-lid dimer interface (28, 30-32). A classic Ser-His-Asp catalytic triad is positioned at the center of a narrow tunnel formed by the lid. The tunnel architecture with open ends is fixed by the dimer interface. In contrast to these dimeric offloading TEs, many PKS and NRPS pathways also have a second monomeric thioesterase called a type II TE, which performs an editing function within the pathway. TE IIs, as well as NRPS offloading TEs, are monomers with a flexible lid domain that appears to control access to the active site (33, 34, 36, 37, 39). The curacin TE sequence has low similarity to sequences in all parts of the TE phylogenic tree (77), lacks an N-terminal extension for dimerization, and has a longer internal lid than other PKS offloading TEs.

To gain further insights and enable mechanistic studies of the novel decarboxylation and sulfate elimination, we report here the crystal structure of CurM TE. The structure of the TE lid and an unusual dimer interface appear to fix the active site in a perpetually open state. A model for β-sulfate recognition was tested by site-directed mutagenesis. The similarity of CurM to other conserved ACP-ST-TE tridomain sequences strongly suggests that CurM TE is part of a new sub-family of thioesterases.

Experimental Procedures

Cloning and site-directed mutagenesis

A construct encoding the TE (CurM residues 1929-2211) was amplified from the cosmid pLM14 (27) and was inserted into pMocr, a vector encoding the fusion protein 6xHis-Mocr for enhanced solubility (78). Site directed mutagenesis was performed using the QuikChange protocol (Stratagene). All constructs were verified by sequencing. The CurM ACP and ST expression plasmids were constructed by Dr. Liangcai Gu and previously described (27).

Protein expression and purification

E. coli strain BL21(DE3) was transformed with expression plasmid, grown at 37°C in 500 ml TB with 4% glycerol to an OD_{600} of 1.0, cooled to 18°C, induced with IPTG (final concentration 0.2 mM), and grown for an additional 18 h. Selenomethionyl (SeMet) protein was produced in the same *E. coli* strain in SelenoMet Medium (AthenaES) containing 100 μg/ml of seleno-DL-methionine.

All protein purification steps were performed at 4°C. The cell pellet from 500 ml of cell culture was resuspended in 40 ml Buffer A (20 mM Tris pH 7.9, 500 mM NaCl, 20 mM imidazole, 10% glycerol), incubated 30 min with DNase (2 mg), lysozyme (5 mg), and $MgCl_2$ (4 mM) and lysed by sonication. The soluble fraction was loaded onto a 5-ml HisTrap Ni NTA column (GE Healthcare). CurM TE was eluted with a linear gradient from 20 mM to 650 mM imidazole (Buffer B). The 6xHis-Mocr fusion partner was removed by 2-h incubation with 1 mM DTT and tobacco etch virus (TEV) protease (1 mg protease per 50 mg TE) at room temperature. After overnight dialysis at 4°C in Buffer C (20 mM Tris pH 7.9, 500 mM NaCl, 10% glycerol) with 1 mM DTT, remaining His-tagged proteins were removed by Ni-affinity chromatography, followed by size exclusion chromatography with a HiLoad 16/60 Superdex 200 column (GE Healthcare) pre-equilibrated with Buffer C. CurM TE was concentrated to 5 mg/ml, flash frozen in liquid N_2, and stored at -80°C. Of 14 TE variants purified as the wild type, only 6 yielded enough soluble protein for assay. SeMet TE was purified as the wild type with addition of 2 mM DTT to all buffers. Yields per 500 ml culture were 5 mg TE and 2 mg SeMet TE.

Crystallization.

Crystals grew at 4°C within 24-48 h by vapor diffusion from a 1:1 mix of protein stock (2 mg/ml TE, 20 mM Tris pH 7.9, 200 mM NaCl, 2.5% glycerol) and well solution (27-32% PEG3350, 100 mM Tris pH 8.3-8.5). Micro-seeding was required for crystal growth of the SeMet protein in similar conditions with 1 mM DTT in the protein solution. Crystals were cryo protected in well solution with 15% glycerol, harvested in loops and flash cooled in liquid N_2.

Data collection and structure determination.

Data were collected at GM/CA beamline 23ID-D at the Advanced Photon Source (APS) at Argonne National Lab (Argonne, IL) (Table 2.1). Among 25 SeMet TE crystals, only one diffracted beyond 4 Å, but was visibly two crystals and had multiple lattices in the diffraction pattern. A region visually identified as a single crystal was probed in three 10-µm steps using a 20-µm mini-beam (79). The crystal was centered at the position with the strongest diffraction and the least interference from the second lattice, and Friedel data were collected in inverse-beam geometry (ϕ = 0°-90° and 180°-270° as wedges of 45° with 1° images) at the wavelength of peak absorption at the Se edge. The diffraction images showed signs of decay at the end of the collection. The crystal was translated to expose a fresh region to the beam, and rotated 90° from the start of data collection. Each of two regions was probed with the 10-µm mini-beam in a 3 x 3 raster of 10-µm steps. Data were collected at the strongest-diffracting position with a single lattice, again in inverse beam geometry (ϕ = 90°-150° and 270° 330° as 30° wedges with 0.5° images). The two partial datasets were integrated separately and scaled together using the HKL2000 suite (80) (Table 2.2). Using Phaser (81) in the PHENIX (82) software suite, Se sites were found for all Met residues of the four polypeptides in the asymmetric unit (Table 2.3). Nine Met side chains had two partially occupied Se sites and another Met had three sites for 39 total Se sites (average figure of merit (FOM) = 0.401). After density modification and fourfold noncrystallographic symmetry averaging in RESOLVE (83) (FOM = 0.81), an 86% complete initial model was built by AUTOBUILD (84) and completed manually in COOT (85). A 1.7 Å native dataset was used for refinement (Table 2.4). REFMAC5 (86), from the CCP4 suite (87), was used for refinement with 5 TLS groups per monomer (87, 86, 88) (Figure 2.2). NCS was not used

during any stage of the refinement. Electron density was complete throughout the polypeptide chain except for two loop regions, which had different disordered residues in the four polypeptide chains (monomer A: 132-134, B: 205-206, C: 133-136, 205-216, D: 206-207). While no single monomer is a complete view of these loops, superposition of the monomers provides a complete model. The atomic coordinates and structure factors (code 3QIT) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org).

Table 2.1 Crystallographic Summary

	CurM TE (SeMet)	CurM TE (native)
Diffraction Data		
Space group	$P2_1$	$P2_1$
X-ray source	APS 23-ID-D	APS 23-ID-D
a, b, c (Å)	74.1, 86.9, 87.1	74.5, 86.9, 87.6
α, β, γ (°)	90, 90.4, 90	90, 90.8, 90
Wavelength (Å)	0.97948	1.0332
d_{min} (Å)	2.14 (2.22-2.14)[a]	1.68 (1.74-1.68)
Avg I/σ_I	16.3 (5.5)	18.1 (2.1)
R_{symm}[b]	0.113 (0.362)	0.059 (0.384)
Completeness	99.5 (99.9)	90.2 (49.8)
Avg. redundancy	6.4 (6.3)	3.4 (2.0)
Unique reflections	60,147	114,584
Refinement		
Data range (Å)		39.10-1.68
No. reflections		108,718
R_{work}/R_{free}[c]		0.164/0.209
RMS deviations		
Bonds (Å)		0.011
Angles (°)		1.268
Avg *B*-factors (Å^2)		
Protein		29.5
Water		39.9
Ramachandran		
Allowed		99.9%
Outliers		0.1%
Protein Atoms		8817
Water Molecules		1203

[a]Outermost shell in parentheses.
[b]Including anomalous differences.
[c]The R_{free} data set included a random 5% of reflections.

Table 2.2 Scaling statistics for SeMet CurM TE

Shell Limit (Å)						
Lower	Upper	Average I	Avg. error	Chi^2	Linear R-factor[a]	Square R-factor
50.00	4.61	118.6	4.9	1.878	0.069	0.068
4.61	3.66	139.2	5.8	1.386	0.073	0.080
3.66	3.20	91.2	4.2	1.191	0.079	0.081
3.20	2.90	50.7	3.0	1.242	0.114	0.113
2.90	2.70	35.2	2.6	1.170	0.143	0.140
2.70	2.54	27.6	2.4	1.073	0.170	0.161
2.54	2.41	22.9	2.4	1.066	0.206	0.202
2.41	2.31	20.8	2.5	1.064	0.240	0.233
2.31	2.22	17.8	2.5	1.047	0.282	0.278
2.22	2.14	14.3	2.6	1.047	0.362	0.348
All	Reflections:	53.8	3.3	1.219	0.113	0.092

[a]Includes anomalous differences

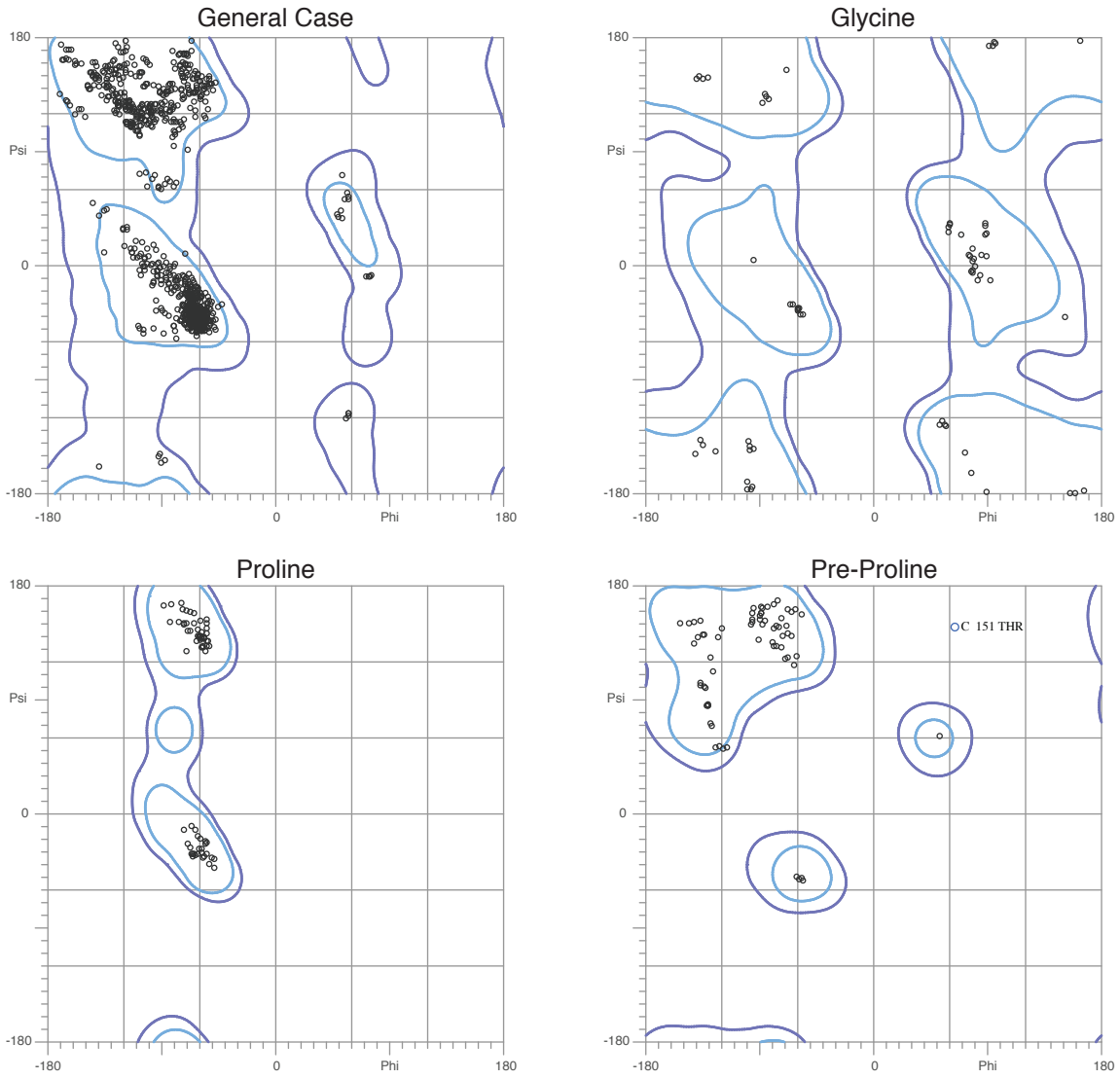
Table 2.3 Phasing statistics for SeMet CurM TE

Anomalous signal	0.094
Sites found	39
NCS copies	4
Sites in each NCS group	6
NCS correlations	0.60
Bayes-CC[a]	48.9
FOM[b]	0.4
LLG[c]	-671,100

[a]Bayesian estimate of the quality of experimental electron density maps.
[b]Final figure of merit of phasing
[c]Log-likelihood gain

Table 2.4 Scaling statistics for Native CurM TE

Shell Limit (Å)						
Lower	Upper	Average I	Avg. error	Chi^2	Linear R-factor	Square R-factor
50.00	3.62	6261.4	277.5	0.997	0.048	0.058
3.62	2.87	3024.2	145.0	0.903	0.051	0.063
2.87	2.51	1152.3	64.1	0.961	0.060	0.066
2.51	2.28	732.0	43.4	1.047	0.072	0.078
2.28	2.12	503.2	38.1	1.075	0.093	0.095
2.12	1.99	320.5	34.2	1.025	0.131	0.128
1.99	1.89	197.9	29.6	1.091	0.176	0.171
1.89	1.81	120.1	28.3	1.156	0.219	0.203
1.81	1.74	72.3	24.1	1.178	0.296	0.265
1.74	1.68	49.4	23.1	1.280	0.384	0.338
All	Reflections:	1366.4	75.6	1.033	0.059	0.06



97.7% (1117/1143) of all residues were in favored (98%) regions.
99.9% (1142/1143) of all residues were in allowed (>99.8%) regions.

Figure 2.2 Ramachandran Analysis of CurM TE
Ramachandran plots of final refined model or CurM TE (PDB code 3QIT). Plots were generated using MolProbity (http://kinemage.biochem.duke.edu) (89).

Sequence alignment, structure alignment and substrate modeling.

The search for ACP-ST-TE homologs was done with BLAST (90), MUSCLE (91) was used for multiple sequence alignment, and PyMOL was used to align structures and prepare figures (92). CurM TE was aligned with affinity-labeled Pik TE (PDB code 2H7X, RMS = 3.309) by superposition of the core domains (residues 55-176 and 232-292

in Pik TE to residues 1-126 and 217-282 in CurM TE). The PRODRG2 server (93) was used to generate coordinates and a topology file for modeling the acyl-enzyme intermediate, which was modeled manually in COOT using the affinity label in the active site of PikTE (30, 31) as a guide.

Enzyme assay.

CurM TE activity was assayed using a modification of our previous protocol (27). Apo-ACP was loaded with a substrate analog by 2-h incubation of 50 μM apo-ACP, 100 μM (3*R*)-hydroxy-5-methoxytetradecanoyl-CoA (27), 10 μM *S. verticillus* Svp (94), 10 mM $MgCl_2$, 100 mM Tris pH 7.9 at 30°C. Complete loading was confirmed 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 30% to 90% CH_3CN (0.1% CF_3CO_2H) / H_2O (0.1% CF_3CO_2H) over 45 min. After exchange into Buffer C and concentration (Amicon Ultra 10 kDa concentrators Millipore), substrate-loaded ACP was flash frozen and stored at -80°C. To generate the sulfated substrate for the TE assay, 225 μM loaded ACP was incubated with 5 μM ST, 1.75 mM PAPS (Sigma), 100 mM Tris pH7.9 at room temperature for 10 min. Complete sulfonation was confirmed by HPLC. The TE reaction was initiated by addition of TE (4 μM). After 1 min the reaction was quenched with 10% formic acid. Conversion of loaded to holo ACP was quantified by HPLC as described above. Assays with the non-sulfated substrate were performed by incubating 1 mM (3*R*)-hydroxy-5-methoxytetradecanoyl-CoA, 50 mM Tris pH7.9, with 40 μM TE for 16 h, including 6% glycerol and 300 mM NaCl for protein stability. The reaction was quenched with equal volume of 1M CH_3CO_2H and neutralized with 1M NaOH, and crotonyl-CoA was added as an internal standard. Hydrolysis was analyzed using a Luna C18 column (250 x 4.60 mm, 5 μm, 100 Å, Phenomenex) with a linear gradient from 10% to 90% CH_3OH/H_2O (10 mM $CH_3CO_2NH_4$) over 20 min.

Results

Overall structure.

The curacin TE is the C-terminal domain of the 2211-residue CurM polypeptide. Because low sequence identity to other TEs prevented accurate definition of the domain N-terminus, expression plasmids were constructed with three different start sites corresponding to amino acids 1917, 1929, and 1934. For all three domain variants, an N-

terminal fusion of the protein Mocr (78) was necessary to obtain sufficient soluble protein for purification and crystallization. The construct encoding amino acids 1929-2211 of CurM (here denoted 1-283) yielded crystals after removal of the Mocr fusion partner. In the CurM TE crystal structure, solved by SAD phasing using selenomethionyl CurM TE (Table 2.1), the polypeptide chain was ordered to the ends of the construct.

CurM TE possesses the α/β hydrolase fold, as expected, with residues 1-132 and 215-283 comprising the structurally conserved core domain (Figure 2.3A,B). Residues of the catalytic triad (Ser100, Glu124 and His266) are located at the top of the core domain as in other TEs, but with a Glu in place of the more common Asp (Figure 2.3B). The catalytic triad faces into a cleft between the core and lid. The lid sub-domain (residues 133-214) is ~20 residues longer than the analogous region in other TEs, and includes three helices (αL1, αL2, αL3) and a small β-hairpin (βL1, βL2) (Figure 2.3A,B). The first lid helix (αL1) is designated the "protruding helix" because it has few contacts with the lid and none with the core of the polypeptide.

The orientation of the lid with respect to the core is identical in the four independent copies of the CurM TE polypeptide in the asymmetric unit of crystals (rmsd = 0.29 Å for 216 Cα atoms, Figure 2.4A), demonstrating that the active site cleft is identically open in all four polypeptides. The position of the lid is maintained by complementary surface contacts of lid helix αL3 (residues 176-186) with several loops in the core. Most of the contacts are hydrophobic. Specificity for the fixed lid-core orientation is provided by an "arginine anchor" in which the side chain of Arg185 in lid helix αL3 extends into the core domain where it forms a full set of five hydrogen bonds with core residues, including a buried salt bridge with Asp57 and hydrogen bonds with Gln35 and Glu3 (Figure 2.5). Additionally, the Gln35 side-chain amide is hydrogen bonded with the backbone carbonyl of Leu182 in lid helix αL3. In contrast to the remarkable lid-core surface complementarity on one side of the active site cleft, the linker peptides (residues 129-136 and 205-215) on the opposite side of the cleft are dynamic with some residues disordered in some subunits (Figure 2.4A).

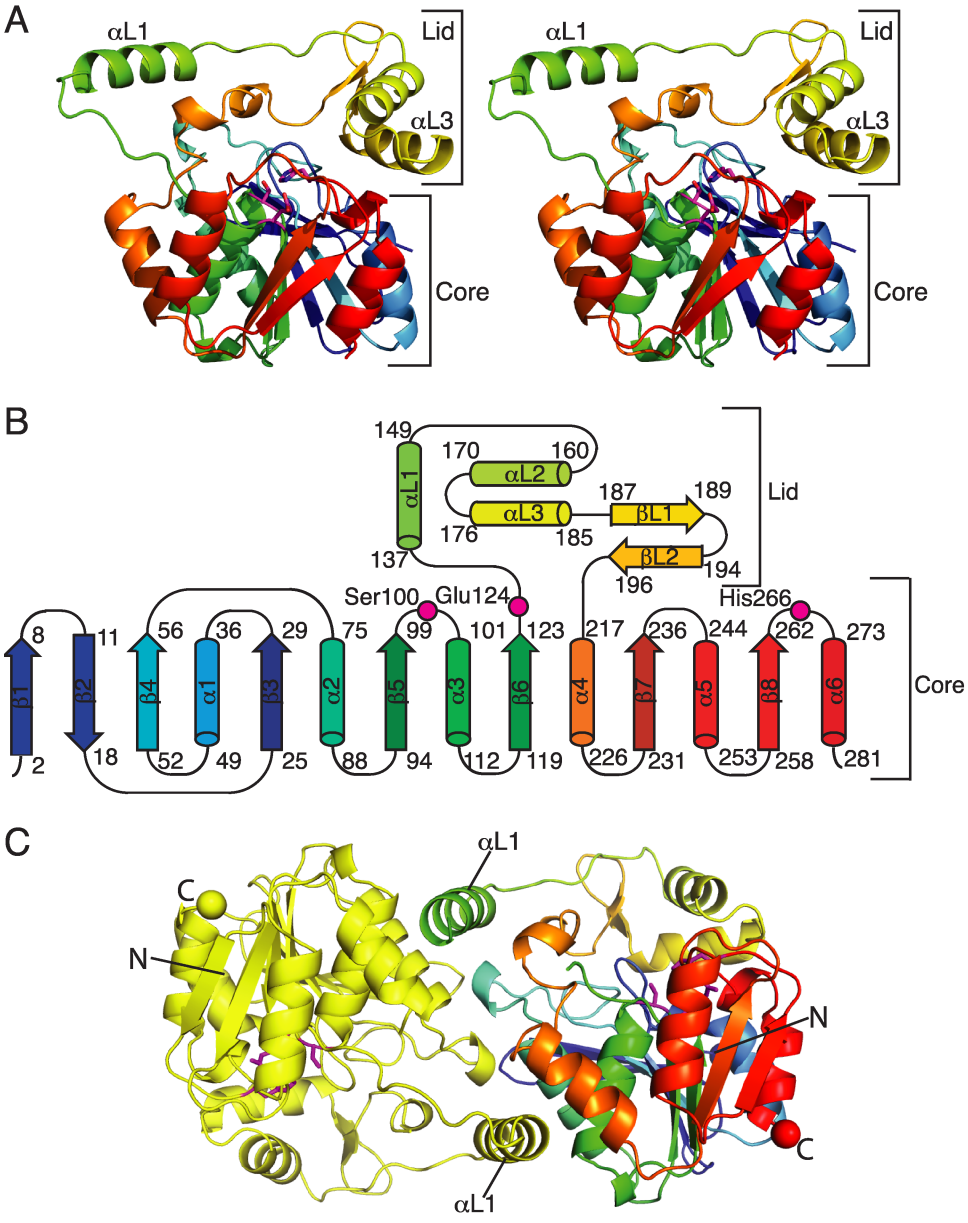


Figure 2.3 Structure of curacin A thioesterase

A. CurM TE polypeptide. The stereo ribbon diagram is colored as a rainbow from blue at the N-terminus to red at the C-terminus with the catalytic triad residues in stick form with magenta C. **B.** Topology diagram. CurM TE has an α/β hydrolase fold in the core domain and a novel lid topology. Residues of the catalytic triad (Ser100, Glu124, His266) are labeled. **C.** Backbone trace of the CurM TE dimer viewed along the molecular dyad. Monomers are colored as a rainbow (right) and in yellow (left), with the catalytic triad as in 2A, and N- and C-termini shown as spheres of the same color as the terminal residue.

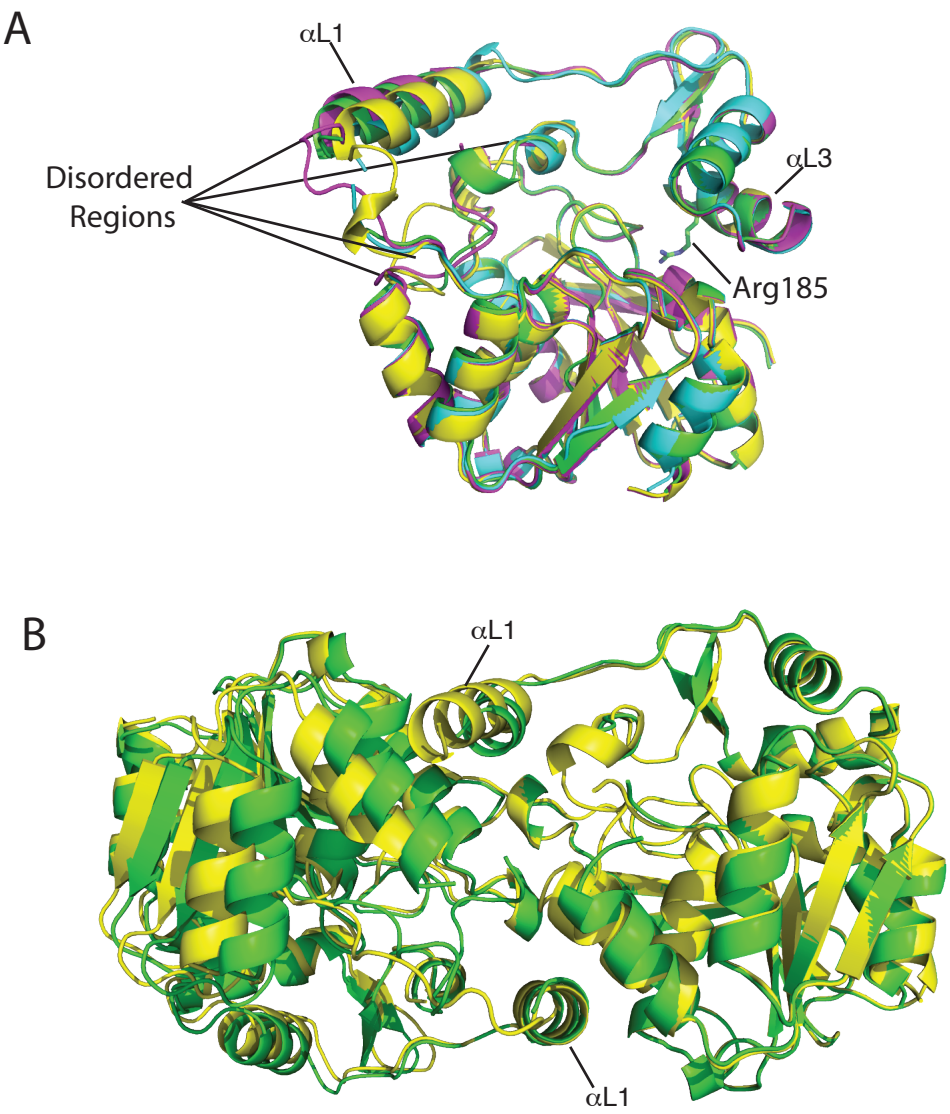


Figure 2.4 Structural identity of the four independent polypeptides and two dimers in CurM TE crystals

A. Superposition of the four monomers in the crystallographic asymmetric unit (rmsd = 0.285 Å for 216 Cα atoms). Slight differences exist in the protruding helix, αL1, which participates in the dimer interface. Some regions of the lid-to-core linkers are disordered in some monomers. View is like that in Figure 2.3A. **B**. Structure alignment of the CurM TE dimer based on superposition of one monomer. A 3º flexing of the dimer interface is apparent in this superposition of the subunits on the right. The protruding lid helix, αL1, clearly moves with the partner subunit in dimer flexure.

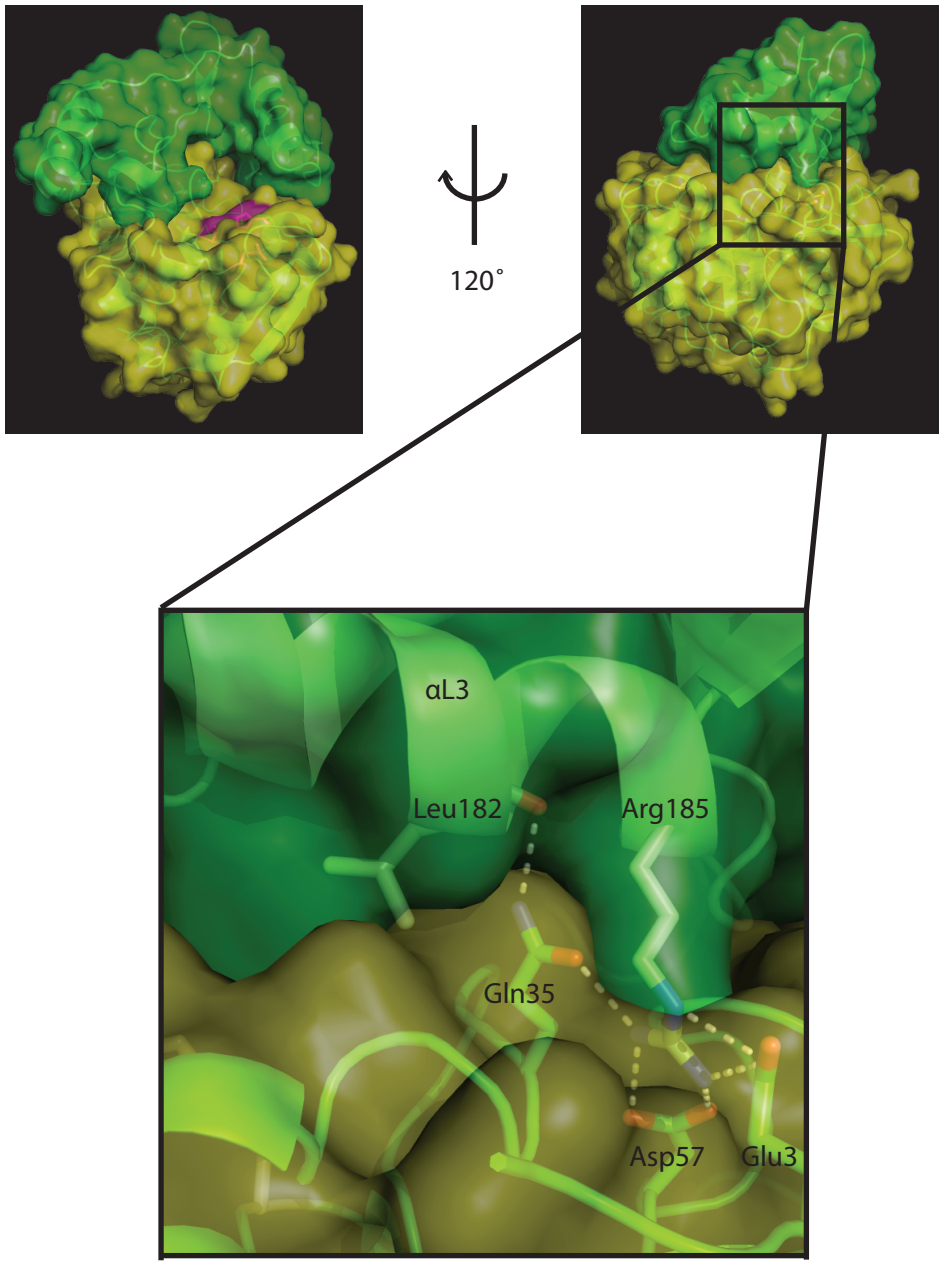


Figure 2.5 Lid-to-core interface
Interactions of lid helix αL3 with the core. Molecule is colored as in Figure 3B: Core is yellow, lid is green and active site residues are magenta. Arg185 forms a buried salt bridge with Asp57 and hydrogen bonds with Gln35 and Glu3. The Gln35 side-chain amide is hydrogen bonded with the backbone carbonyl of Leu182 in lid helix αL3.

Novel dimer interface.

CurM TE is dimeric in solution (Figure 2.6) and also in the crystal structure (Figure 2.3C). The mode of dimer formation is radically different than in other PKS offloading TEs even though the core structures and catalytic triads are similar (Figure 2.7A). The primary dimer contact is between the protruding helix (αL1) in the lid of one subunit and three helices (α2, α3, α4) in the core of the partner subunit (Figure 2.3C), with additional core-to-core contacts of the β4-α2 loops of the two subunits. The subunit interface is predominantly hydrophobic and large (buried surface area of 1220 $Å^2$ per monomer). The N-termini are distant from the dimer interface (Figure 2.3), allowing for fusion to the monomeric ST domain as well as dimerization of the TE domain within the dimeric CurM module. This is consistent with the observation that the natural CurM ST-TE di-domain is dimeric in solution (data not shown).

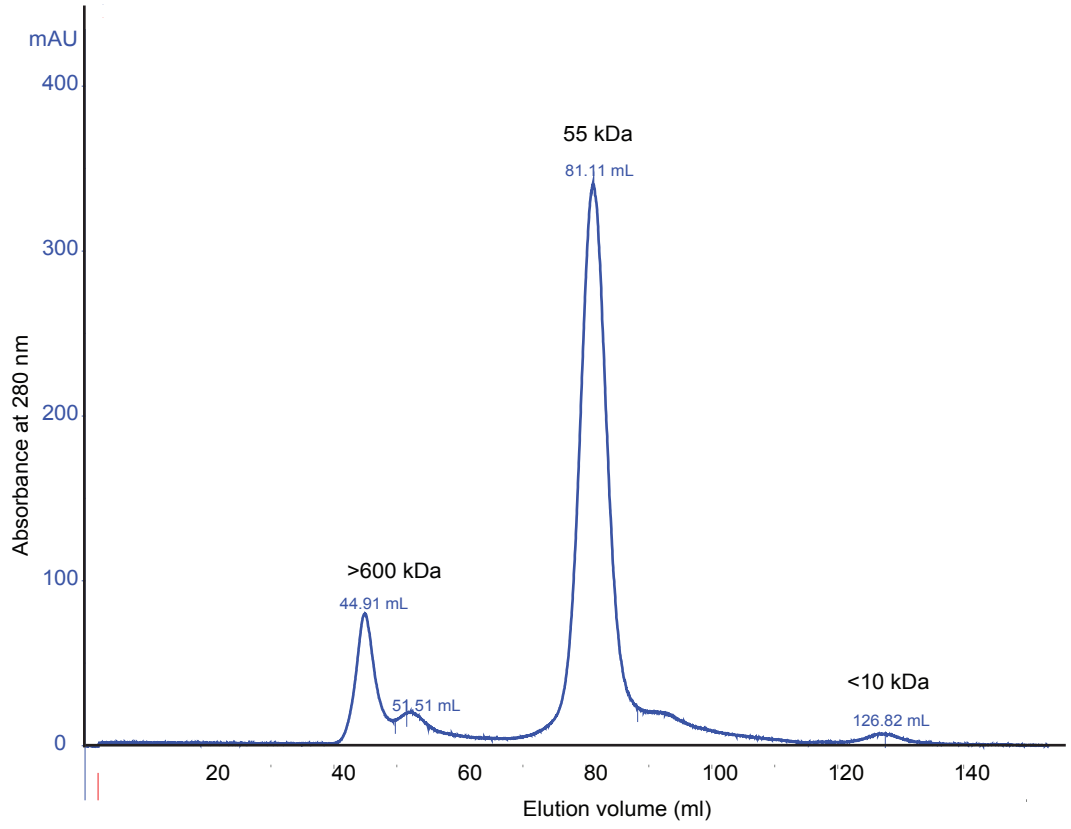


Figure 2.6 Size exclusion analysis of CurM TE
In the elution profile from a HiLoad 16/60 Superdex 200 size exclusion column, CurM TE has an apparent molecular mass of 55 kDa, similar to the calculated molecular weight of a CurM TE dimer (62.2 kDa).

The crystal asymmetric unit contains two dimers that differ by a slight flexure (3º) at the dimer interface, accounting for the poorer overall fit of dimers (rmsd = 1.05 Å for 486 Cα atoms) compared with monomers. The protruding lid helix, αL1, moves with the partner subunit as the subunits flex (Figure 2.4B). Thus, the lid-to-core dimer contact helps fix the lid in an open orientation (Figure 2.7B). This is surprising and unique

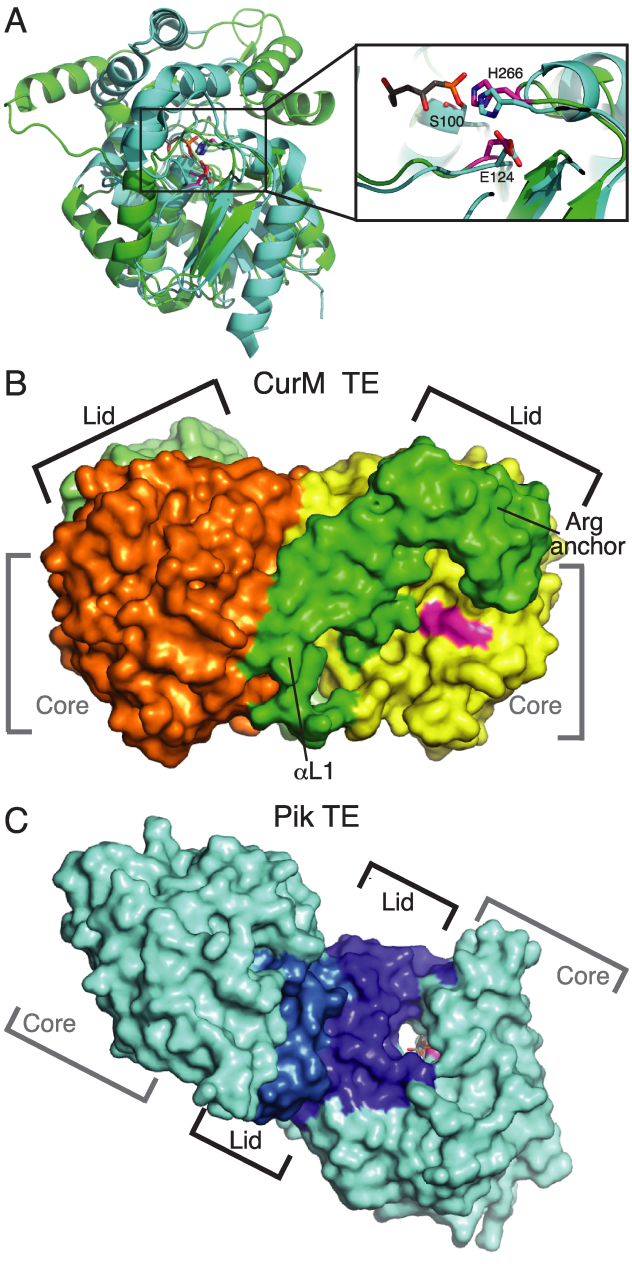


Figure 2.7 Comparison of curacin and pikromycin TEs
A. Structure alignment of the core of CurM TE (green) and Pik TE (cyan, PDB 2H7X, (12)) (RMSD = 1.5 Å for 95 Cα atoms). Both structures have the conserved α/β hydrolase core, but the lids differ. The zoom view shows the active-site conservation in the catalytic triad of CurM TE (magenta) and Pik TE (cyan) with a triketide affinity label (gray). The view is similar to Figure 2.3A. **B.** Surface representation of the CurM TE dimer. The primary dimer contact is between lid (subunits in two shades of green) and core (subunits in yellow and orange). The active site (magenta) is in an open cleft between core and lid. **C.** Surface representation of the Pik TE dimer. The dimer contact is exclusively between the lid subdomains (subunits in two shades of blue) with no contacts of core domains (cyan). The active site (magenta) with an affinity label (gray sticks) is at the center of an open-ended tunnel ((12)). The views in B and C highlight the differences in active site access, which are distinct for the CurM TE and Pik TE enzymes.

compared to the offloading TEs of other PKSs in which lid-to-lid dimer contacts fix the lid in a closed orientation. In these other TEs, the lid-to-lid dimer contact stabilizes an open-ended tunnel with the active site at its center (Figure 2.7C). In contrast, the dimer-enforced open lid of CurM TE results in a highly exposed, open-cleft active site (Figure 2.7B). Thus, the dimer interface creates a very different active site environment in CurM TE and the other offloading PKS TEs.

CurM TE is not alone

We identified five open reading frames in the sequence database with substantial identity (33%-51%) to the CurM ACP-ST-TE tridomain. The sequences are uncharacterized protein products from genomes of cyanobacteria (*Synechococcus* and *Cyanothece*) and proteobacteria (*Pseudomonas* and *Haliangium*). Based on the conservation of all three domains, including for example the TE catalytic triad, we assume that the other gene products also catalyze sequential sulfonation, hydrolysis, decarboxylation, and sulfate-elimination reactions and that conserved residues within these sequences may illuminate areas of the structure that are important for function. Conserved residues in CurM TE and the five other putative TE sequences (Figure 2.8) were mapped onto the CurM TE structure, revealing a dense area of conservation in the active site cleft (Figure 2.9). Sequence conservation indicates that CurM TE and the other gene products have the same lid structure and orientation and the same dimer interface. The essential features of the lid-to-core arginine anchor (Arg185, Asp57, Gln35) are conserved. The protruding helix (αL1) has conserved hydrophobic character, as do the surfaces of the core that it contacts in the dimer. The lid-to-core TE dimer

interface, the open active site cleft, and the surface of the cleft are likely conserved among these TEs.

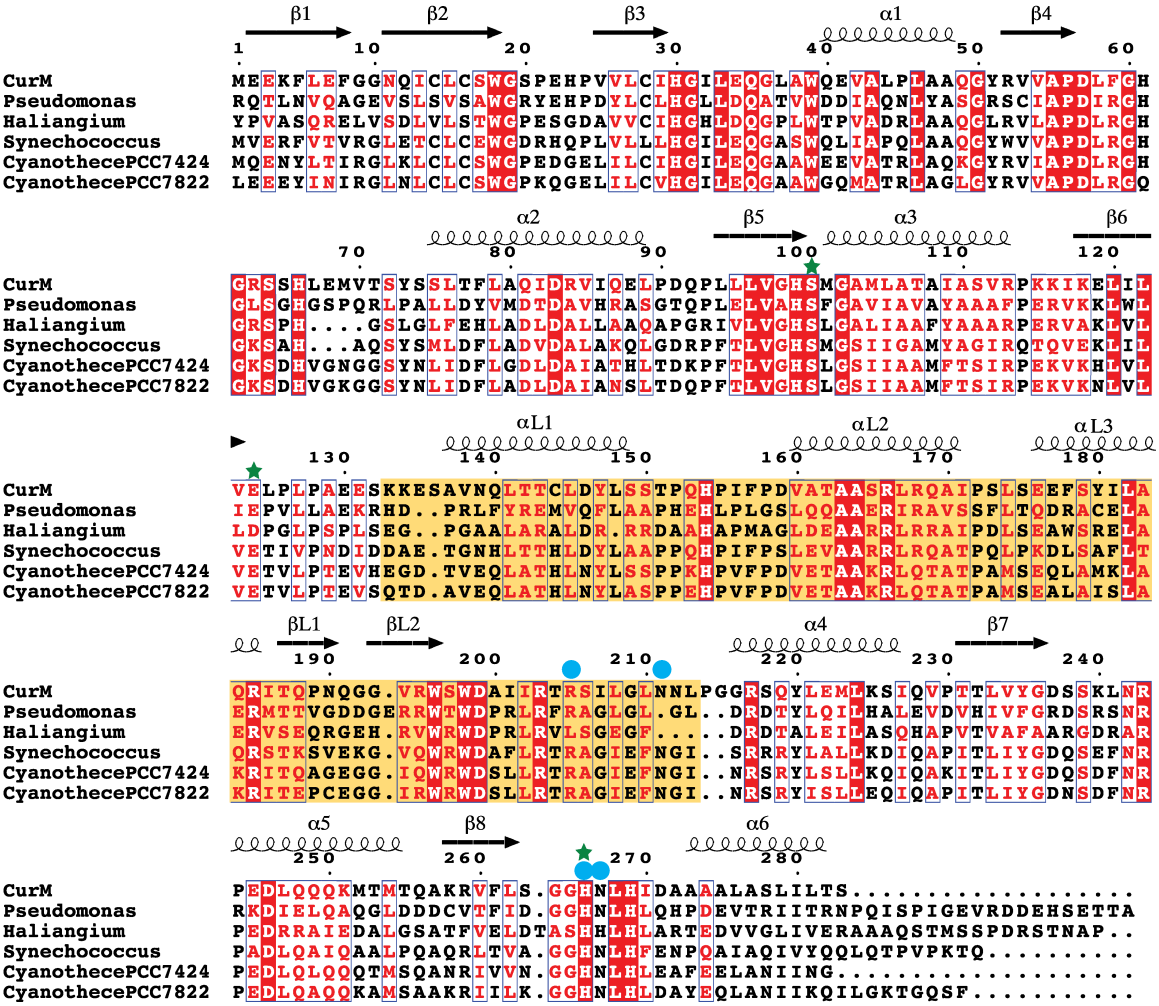


Figure 2.8 Sequence alignment of CurM TE with putative TEs from open reading frames encoding tandem ACP-ST-TE tridomains

Invariant residues are highlighted in red, conserved residues are printed in red. The conserved lid domain is highlighted in yellow. Green stars indicate active site residues and blue circles indicate residues probed by site-directed mutagenesis. Sequence alignment was performed by MUSCLE (91), and the figure was prepared using ESPript (95). GenBank entries are: *Pseudomonas entomophila* L48 (GenBank YP_610919), *Haliangium ochraceum* DSM 14365 (YP_003265308), *Synechococcus* PCC 7002 (YP_001734428), *Cyanothece* PCC 7424 (YP_002377174), and *Cyanothece* PCC 7822 (ZP_03153601).

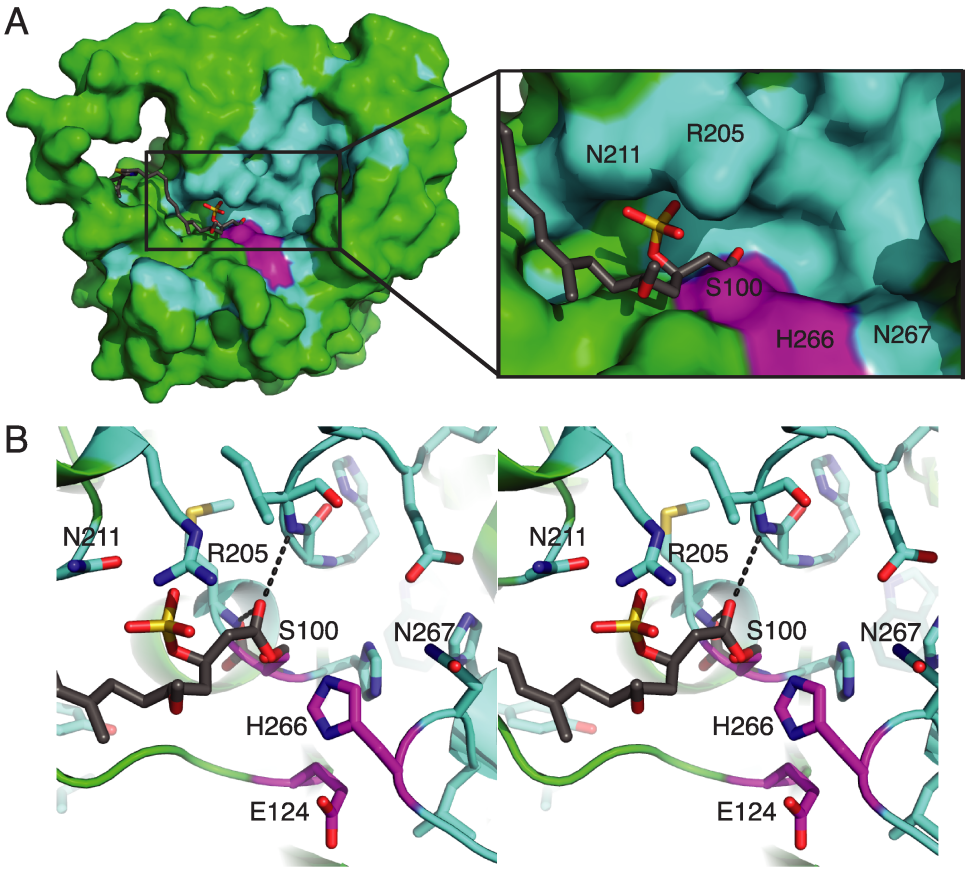


Figure 2.9 CurM TE active site
A. Surface of the active-site cleft. In the molecular surface view, the active site (magenta catalytic triad) is in a cleft in which the presumed phosphopantetheine entrance is lined with conserved residues (cyan), with other parts of the protein surface in green and the modeled acyl-enzyme intermediate in gray sticks. **B.** Modeled acyl-enzyme intermediate (gray C). In this stereo view, the intermediate is surrounded by conserved amino acid residues (cyan C) and the catalytic triad (Ser100, Glu124 and His266) (magenta C). The carbonyl oxygen is bound in the oxyanion hole (hydrogen bonds to the NHs of the Ile32 and Met101).

Active site

The open-cleft active site and intact catalytic triad (Figure 2.7) appear poised to hydrolyze a broad array of thioester substrates. To understand how CurM TE selects a sulfated substrate, we modeled the acyl-enzyme intermediate of the hydrolytic reaction (Figure 2.10) using our knowledge of catalytic triad catalysis and geometry and the structure of affinity-labeled Pik TE (31) (Figure 2.9B). Constraints on the model included acylation of the Ser100 nucleophile, carbonyl-oxygen binding in the oxyanion

hole (NHs of Ile32 and Met101), the (*R*)-β-sulfate isomer, contact of sulfate with the enzyme, and placement of the long acyl chain away from the catalytic His266. Given these constraints, the acyl enzyme was easily modeled with favorable bond rotations and without change to the protein structure. The acyl chain lies in a narrow cleft between the lid and core domains, and the (*R*)-β-sulfate occupies a niche adjacent to Arg205 (Figure 2.9A). The model is consistent with the strong stereoselectivity of CurM TE for the (*R*)-β-sulfate over the corresponding (*S*) isomer (27), as modeling of the (*S*)-β-sulfate form of the substrate resulted in sulfate contacts with the hydrophobic surface of the protein. The model also suggests that the hydrogen-bonded side chains of Arg205 and Asn211 may recognize the β-sulfate moiety. To explore these issues further, the functions of the presumed sulfate-recognition residues and others in the active site cleft were probed by site-directed mutagenesis. Three substitutions were made at Arg205 (Gln, Glu and Ala), as well as Ala substitutions at Asn267, which is adjacent to the catalytic His, and at Asn211, which is hydrogen bonded with Arg205 (Figure 2.9B).

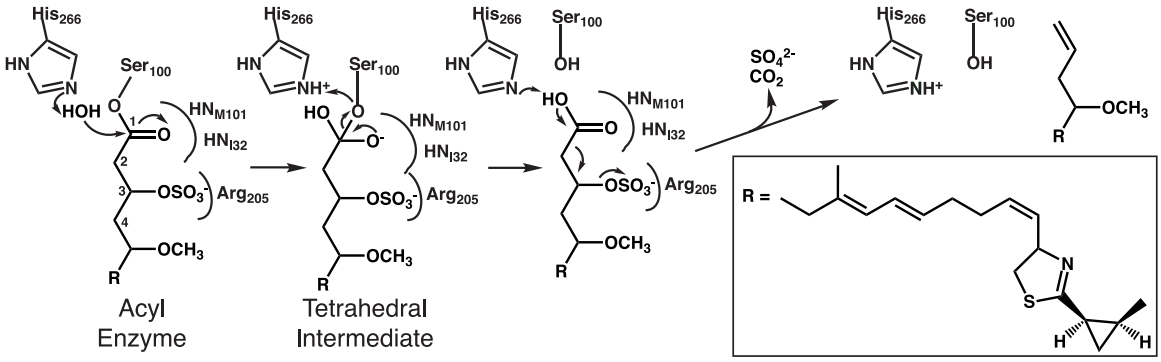


Figure 2.10 Predicted CurM TE reaction sequence
Predicted reaction sequence for CurM TE hydrolysis, sulfate group elimination and decarboxylation. Arg205 recognizes the (*R*)-β-sulfate, facilitating hydrolysis by the catalytic triad. The catalytic His266 is positioned to assist in decarboxylation followed by sulfate elimination.

The TE wild type and variants were tested using a one-pot, multi-step assay (27) in which recombinant CurM ACP was loaded with a synthetic substrate mimic (Figure 2.1B) and reacted sequentially with recombinant CurM ST and CurM TE. ACP substrates and ACP hydrolysis products were analyzed by HPLC (Figure 2.11). Activities of the CurM TE variants were normalized to the activity of the wild type

(Table 2.5). The three substitutions at Arg205 (R205Q, R205E, and R205A) resulted in significantly reduced activity, comparable to a negative-control substitution in the catalytic triad (H266R). Alanine substitutions at two Asn residues that are conserved in the ACP-ST-TE group of sequences but not in the wider TE family (N211A and N267A) resulted in only two- to three-fold reduced catalytic activity. To test the role of Arg205 in sulfate recognition, the Arg205 variants were assayed with a non-sulfated substrate. Due

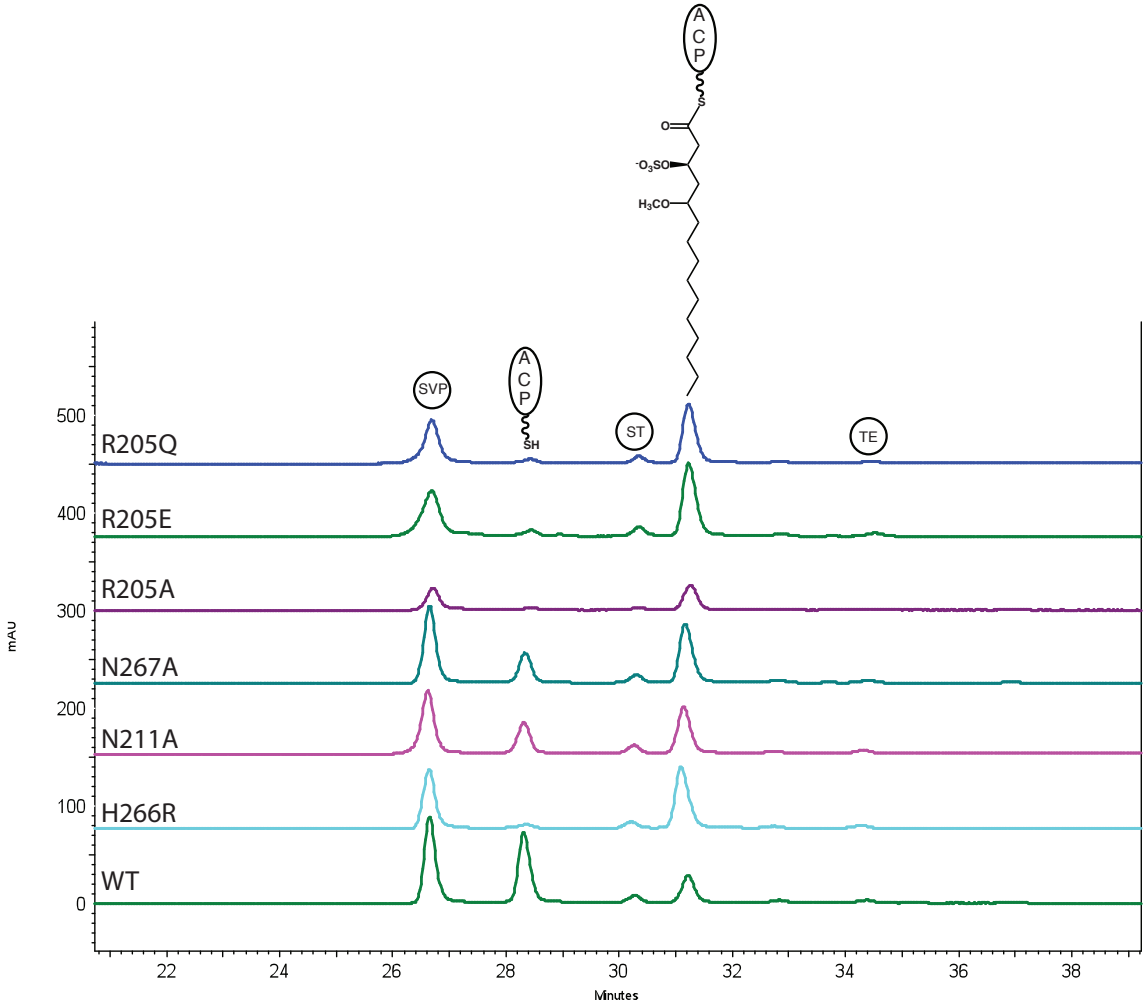


Figure 2.11 HPLC detection of TE substrates and products
The one-pot assay mixture was separated by reverse-phase HPLC resulting in peaks for Svp, CurM ST, and CurM TE as well as the assay substrate, δ-methoxy-β-sulfate acyl loaded CurM ACP, and the product, CurM holo-ACP. In these HPLC chromatograms, the Svp peak area was used as an internal standard to correct the holo-ACP peak areas for differences in injection volume. Results are reported as the fraction of the holo-ACP peak area for the wild type.

to the ~800-fold lower activity of wild type TE on non-sulfated substrates than on sulfated substrates (27), these assays were performed with a CoA-linked β-OH substrate available at higher concentration than the ACP-linked substrate. In striking contrast to their lack of activity with sulfated substrates, the Arg205 variants had activity comparable to the wild type with the non-sulfated substrate (Table 2.6). This result supports a vital role for Arg205 in sulfate recognition.

Table 2.5 Activity of CurM TE variants

	Activity (%)
WT	100
R205Q	1.8±0.3
R205E	2±2
R205A	6±3
N267A	34±3
N211A	55±2
H266R	1.4±0.2

Raw HPLC chromatogram peak area for the holo-ACP product with background subtracted and normalized with an internal standard to correct for differences in sample injection. The area for each mutant was normalized to the wild type within each replicate. Means ± standard deviations from duplicate experiments are shown.

Table 2.6 TE activity with non-sulfated CoA substrate

	Activity (%)
WT	100
R205Q	91±4
R205E	90±9
R205A	93±6

The non-sulfated CoA substrate, (3*R*)-hydroxy-5-methoxytetradecanoyl-CoA, was reacted with CurM TE for 16 h. The reaction was monitored by HPLC analysis of CoA species. The fraction of substrate hydrolyzed was calculated from the ratio of peak areas of the substrate ((3*R*)-hydroxy-5-methoxytetradecanoyl-CoA) and product (CoA). The negative-control catalytic-triad substitution (H266R) was used for background subtraction, and results were normalized to the wild type within each replicate. Means ± standard deviations of duplicate experiments are shown.

Discussion

A new thioesterase sub-family.

The CurM TE structure defines a new thioesterase sub-family with the remarkable catalytic activity of hydrolysis, decarboxylation, and sulfate elimination to form alkene products from β-sulfated acyl-ACP substrates. The structural features that distinguish CurM TE from other thioesterase sub-families are conserved in sequences of five other putative ACP-ST-TE tridomains, indicating that all of them catalyze terminal alkene formation. The CurM TE lid sub-domain, which sits above the active site, is substantially longer than the analogous region in other TE sub-families. The lid creates an open active site cleft, and is fixed in an open orientation by key interactions unique to this TE sub-family (Figure 2.3C, Figure 2.7B). One side of the lid is secured by the surface complementarity of lid helix αL3 with the core domain and by an "arginine anchor" in which Arg185 in the lid extends deep into the core domain to form a salt bridge with Asp57. The other side of the lid is secured by dimer contacts of lid helix αL1 with the core domain of the partner subunit.

The characteristics of the CurM TE lid provide a striking contrast to other PKS and NRPS TE sub-families. Dimeric PKS offloading TEs have a closed lid, formed by a lid-to-lid dimer contact, which creates an open-ended tunnel with the active site at its center (Figure 2.7C). Monomeric editing TE II enzymes and NRPS offloading TEs have a flexible lid that may control access of the phosphopantetheine arm to the active site by its ability to open and close. The region of substrate entrance in these other TEs (the active site tunnel in other dimeric TEs and the flexible part of the movable lid of monomeric TEs) is analogous to the region in CurM TE secured by the arginine anchor. Hence, the usual entrance route for acyl-ACP substrates is blocked in CurM TE by the anchoring of lid helix αL3 to the core domain. Instead, we presume that substrate enters the CurM TE active site via the conserved cleft that extends away from the catalytic triad towards the top of the lid (Figure 2.9A). In contrast to the firm anchoring on the entrance side of the CurM TE active site cleft, the opposite side of the cleft has flexible lid-to-core linkers and less conservation.

Substrate specificity.

Despite the departure from precedent in the lid architecture and substrate entrance, the catalytic triad of CurM TE is fully formed and appears to be poised for hydrolysis (Figure 2.7A). The question remains why CurM TE does not hydrolyze any thioester substrate given the positions of the Ser100 nucleophile, His266 base and Ile32/Met101 oxyanion hole. An answer is found in the wide-open architecture of the active site, which provides few obvious substrate recognition elements. CurM TE does not need exquisite recognition of its natural substrate, which is present at high effective concentration via tethering to CurM ACP. However, CurM TE must select the (*R*)-β-sulfated substrate and not other acylated forms of CurM ACP (malonyl, β-keto and (*R*)-β-hydroxy), which are also present at high effective concentration before they are transformed by the other catalytic domains of CurM (ketosynthase, ketoreductase, and sulfotransferase, respectively). Our model of the TE acyl-enzyme intermediate suggested that Arg205 recognizes the (*R*)-β-sulfate moiety (Figure 2.9), and this was consistent with the site-directed mutagenesis data (Table 2.5, Table 2.6). The Arg205 variants lacked activity with a sulfated substrate but had activity comparable to the wild type in the much slower reaction with a non-sulfated substrate. Thus Arg205 may be the only positive substrate-recognition element, the active site may have little or no affinity for the rest of the acyl chain, and non-sulfated substrates may be only rarely positioned properly for catalysis. In this manner, the β-sulfate acts as a "handle" for Arg205 to assist in effective thioester binding to the catalytic machinery in the large open active site cleft, whereas non-sulfated substrates have no such guidance. The geometry of Arg205 with relation to the active site also ensures that (*R*)-β-sulfated substrates can be properly positioned for catalysis whereas steric clashes would occur for the corresponding (*S*)-β-sulfated isomer. Intriguingly, Arg205 is among the flexible residues in the linker connecting the CurM TE core and lid, and thus may provide an early interaction with incoming sulfated substrates. The flexibility of this region is consistent with the lack of conservation of Arg205 in the homolog from *H. ochraceum*, in which nearby arginine side chains may serve the same function.

Catalytic mechanism.

The CurM TE structure reveals an intact hydrolytic active site into which an acyl enzyme intermediate was modeled easily without the need to move interacting groups

(Figure 2.9). This strongly suggests that acyl enzyme formation precedes other catalytic events. Therefore, we propose that hydrolysis of the acyl enzyme precedes decarboxylation and sulfate group elimination, based on our detection of both the terminal alkene product and the substrate analog bearing intact β-sulfate and carboxylic acid groups (*i.e*. the initial TE serine ester hydrolysis product) (27). The CurM TE assures decarboxylation and sulfate group elimination *in vivo*, as the corresponding carboxylated and sulfated curacin A intermediate was not detected in cultures of *M. producens* (Figure 2.10) (3). Through interactions with the oxyanion hole and with Arg205, the carboxylic acid hydrolysis product remains bound to CurM TE long enough for the enzyme to promote loss of CO_2 and SO_4^{2-} in one concerted step. In the reaction scheme, CurM TE assists decarboxylation and sulfate elimination in at least three ways. First, the enzyme binds the carboxylic acid hydrolysis product in a conformation that prevents resonance of COOH with the C_1-C_2 bond to be broken upon decarboxylation. Arg205, through its interaction with the sulfate at C_3, may assist in achieving the optimal conformation. Second, the catalytic His266 likely promotes deprotonation of the carboxylic acid hydrolysis product to facilitate loss of CO_2. Stabilization of the additional negative charge on the departing SO_4^{2-} offered by Arg205 may further reduce energy barriers for initiating this remarkable decarboxylative elimination process.

Potential for hydrocarbon production.

CurM TE is the first member of a branch of the TE family optimized to work in concert with a sulfotransferase to eliminate a carboxyl functional group and create a terminal double bond. The ST-TE system is the first example of a biological sulfonation employed for chemical activation (27). The ACP-ST-TE system exists as an uncharacterized open reading frame in genome sequences of five other bacteria, with a high level of conservation of residues within and surrounding the TE active site and in key contacts between domains and subunits. In CurM and in some ST-TE homologs, the sulfonation-decarboxylation system seems designed to eliminate the terminal carboxylate that would result from canonical TE hydrolysis of an acyl-ACP. Elimination of the terminal carboxyl group is a challenge for efforts to develop liquid biofuels. Thus the ST-TE system is of potential benefit in the development of a direct biosynthetic route to medium and long-chain hydrocarbons from fatty acids. This could be achieved through

engineering a β-hydroxy fatty-acid biosynthetic pathway for offloading using the ST-TE mechanism. The ST-TE route to hydrocarbons is biochemically unique compared to other recent approaches, including a cyanobacterial system using acyl-ACP reductase and aldehyde decarbonylase with fatty-acyl-ACP substrates (96) and an *E. coli* system with fatty-acid biosynthesis engineered for hydrocarbon production (65). Other organisms found to produce hydrocarbons with terminal alkenes, such as *Botryococcus braunii (55)*, probably employ a similar ST-TE offloading strategy. Indeed, *B. braunii* is known to use very long chain fatty acids activated at the β position to produce terminal olefin dienes and trienes through a decarboxylative process (56).

Conclusion.

The curacin TE illustrates the remarkable adaptability of the thioesterase structural framework, here adapted for only those acyl-ACP substrates bearing a sulfate substituent at the β position. The CurM TE is a key component of an unprecedented system for thioester hydrolysis and decarboxylation, working in concert with a fused sulfotransferase to generate a highly favorable leaving group. The CurM TE crystal structure reveals a fully developed hydrolytic active site, an arginine residue for (*R*)-β-sulfate recognition, and an active site cleft that may accommodate a wide variety of acyl substrates.

Author contributions to the manuscript

JGM, LG, JLS, and DHS designed the research. JGM performed the research. PW contributed new reagents (CoA substrates). JGM and JLS analyzed the data. JGM and JLS wrote the paper. JGM, LG, PW, WHG, DHS and JLS revised and edited the paper.

Chapter 3

The Structural Basis of Functional Group Activation by Sulfotransferases in Complex Metabolic Pathways

This chapter is in press in ACS Chemical Biology: McCarthy JG, Eisman EB, Kulkarni S, Gerwick L, Gerwick WH, Wipf P, Sherman DH, Smith JL (2012) The Structural Basis of Functional Group Activation by Sulfotransferases in Complex Metabolic Pathways.

Summary

Sulfated molecules with diverse functions are common in biology, but sulfonation as a method to activate a metabolite for chemical catalysis is rare. Catalytic activity was characterized and crystal structures were determined for two such "activating" sulfotransferases (STs) that sulfonate β-hydroxyacyl thioester substrates. The CurM polyketide synthase (PKS) ST domain from the curacin A biosynthetic pathway of *Moorea producens* and the olefin synthase (OLS) ST from a hydrocarbon-producing system of *Synechococcus* PCC 7002 both occur as a unique acyl carrier protein (ACP), ST and thioesterase (TE) tridomain within a larger polypeptide. During pathway termination, these cyanobacterial systems introduce a terminal double bond into the β-hydroxyacyl-ACP-linked substrate by the combined action of the ST and TE. Under *in vitro* conditions, CurM PKS ST and OLS ST acted on β-hydroxy fatty acyl-ACP substrates; however, OLS ST was not reactive toward analogs of the natural PKS ST substrate bearing a C5-methoxy substituent. The crystal structures of CurM ST and OLS ST revealed that they are members of a distinct protein family relative to other prokaryotic and eukaryotic sulfotransferases. A common binding site for the sulfonate donor 3'-phosphoadenosine-5'-phosphosulfate was visualized in complexes with the product 3'-phosphoadenosine-5'-phosphate. Critical functions for several conserved amino acids in the active site were confirmed by site-directed mutagenesis, including a

proposed glutamate catalytic base. A dynamic active-site flap unique to the "activating" ST family affects substrate selectivity and product formation, based on the activities of chimeras of the PKS and OLS STs with exchanged active-site flaps.

Introduction

Sulfotransfer is a ubiquitous and functionally diverse process across all forms of life. Although the acceptor molecule and the specific biological role of sulfonation varies widely, the sulfotransferases (STs) that catalyze transfer of sulfonate from the biological donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to a hydroxy or amine acceptor are generally highly conserved (44, 97, 47, 98). Sulfotransfer in the curacin A biosynthetic pathway contrasts with all of the other known sulfonation events because the natural product is transiently sulfonated, and the final structure of this potent cancer cell cytotoxin includes no sulfate group (12).

The antimitotic natural product curacin A is produced by the marine cyanobacterium *Moorea producens* [formerly *Lyngbya majuscula (4)*] in a biosynthetic pathway containing a unique ST with low sequence similarity to previously described STs. Located within the final polyketide synthase (PKS) monomodule, CurM, the ST is the penultimate domain between the acyl carrier protein (ACP) and thioesterase (TE), two common domains involved in PKS and fatty acid synthase (FAS) pathway termination (2) (Figure 3.1A). In previous studies, we found that CurM ST transfers sulfonate from the donor PAPS to the β-hydroxy of the pathway intermediate, which is covalently tethered to the CurM ACP (Figure 3.1A) (27). The unique CurM TE domain then acts specifically upon the β-sulfated substrate to hydrolyze, decarboxylate and eliminate sulfate to yield a terminal alkene product (Figure 3.1A) (27, 99). This mode of polyketide chain offloading to a terminal alkene was unprecedented, as typical TEs catalyze hydrolysis to the linear acyl-carboxylic acid or cyclization to a macrolactone or macrolactam product (28-32). CurM ST activates the β-hydroxy group and thus facilitates decarboxylation to yield the terminal olefin (27). To our knowledge, this work provided the first description in a metabolic pathway where sulfonation functions to activate a hydroxy group for subsequent conversion to a final product.

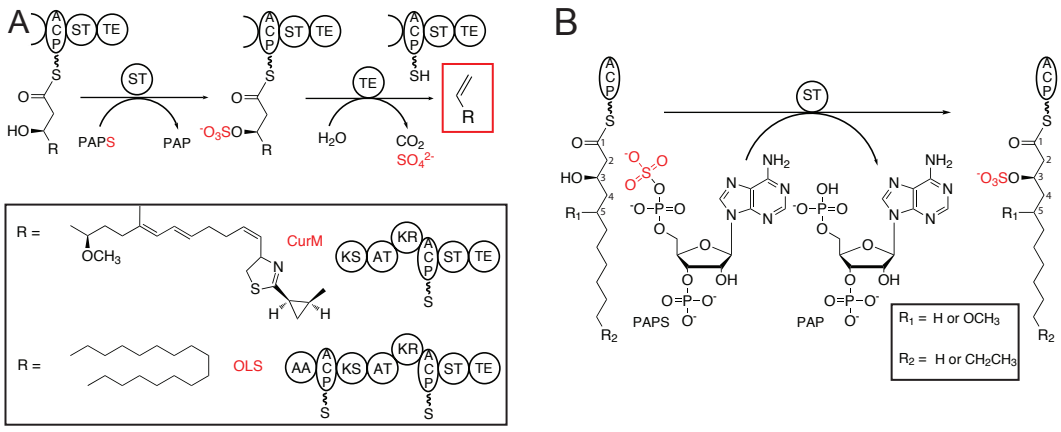


Figure 3.1 Offloading and termination using β-hydroxy group activation and decarboxylation

A. Offloading and termination in the curacin A (CurM) and the olefin synthase (OLS) pathways. The sulfotransferase (ST) first sulfonates the β-hydroxy group from the sulfonate donor 3′-phosphoadenosine 5′-phosphosulfate (PAPS). The thioesterase (TE) then works in a β-sulfate-dependent manner to hydrolyze, decarboxylate, and eliminate sulfate to produce a terminal alkene. Other domains are acyl activating (AA), acyl carrier protein (ACP), ketosynthase (KS), acyl transferase (AT), and ketoreductase (KR). **B.** ST reaction in biochemical assays. Excised ACP was loaded with synthesized substrates (varying in acyl chain length, the presence of a C5-methoxy and the β-hydroxy group configuration) and reacted with PAPS. Sulfonated products were detected by HPLC.

CurM ST has low sequence similarity to all previously characterized STs (less than 25% identity). However, several additional ST domains presumed to play a role in acyl-hydroxy group activation were identified in the protein databases by virtue of a high level of amino acid sequence identity (33–52%) to the CurM ACP-ST-TE tri-domain (Figure 3.2) (10). Here we designate these as "activating" STs with a common functionality to transfer a sulfonate that is subsequently removed through decarboxylative elimination. In addition to *curM*, where the ST is embedded within the CurM PKS module, two related genes appear in the context of uncharacterized biosynthetic gene clusters. Five additional ST genes reside within stand-alone eight-domain polypeptides that resemble monomodular PKSs or FASs. The domain organization is similar to CurM but with the addition of acyl-activating (AA) and ACP domains at the N-terminus (Figure 3.1A, Figure 3.3). The AAs in the stand-alone systems resemble those shown to ligate a

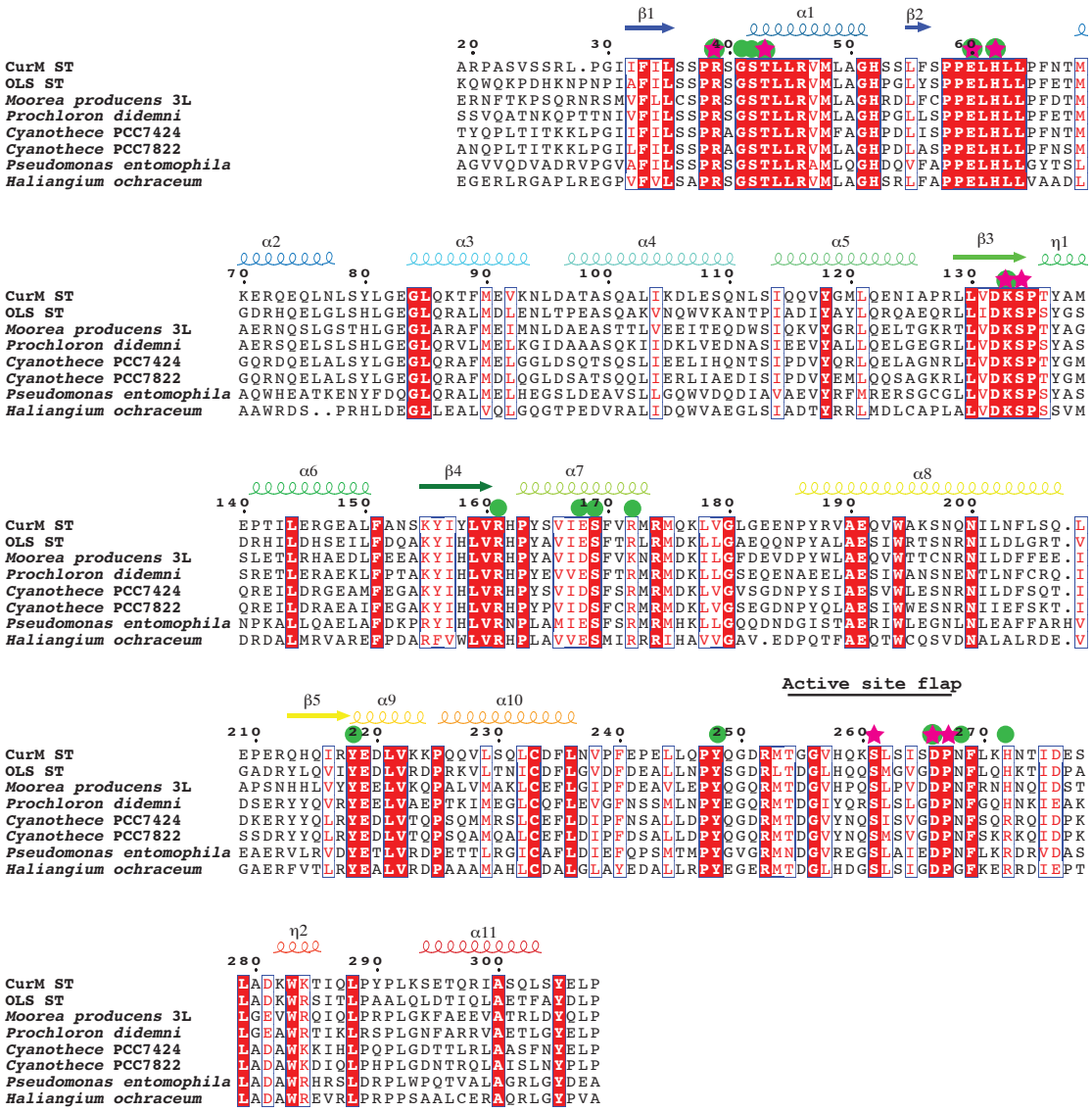


Figure 3.2 Sequence alignment of activating STs

CurM ST and OLS ST are aligned with putative STs from open reading frames containing tandem ACP-ST-TE tridomains. Invariant amino acids are highlighted in red, conserved residues are printed in red. Green circles indicate amino acids in the active site and magenta stars indicate those probed by site-directed mutagenesis. Secondary structures are indicated above the alignment. Sequence alignment was performed by MUSCLE (17) and the figure was prepared using ESPript (24). GenBank entries are: *Moorea producens* (CurM ST) (GenBank ACV42478), *Synechococcus* PCC 7002 (OLS ST) (YP_001734428), *Cyanothece* PCC 7424 (YP_002377174), *Cyanothece* PCC 7822 (ZP_03153601), *Moorea producens* 3L (ZP_08425908), *Prochloron didemni* (AEH57210), *Pseudomonas entomophila* L48 (YP_610919), *Haliangium ochraceum* DSM 14365 (YP_003265308).

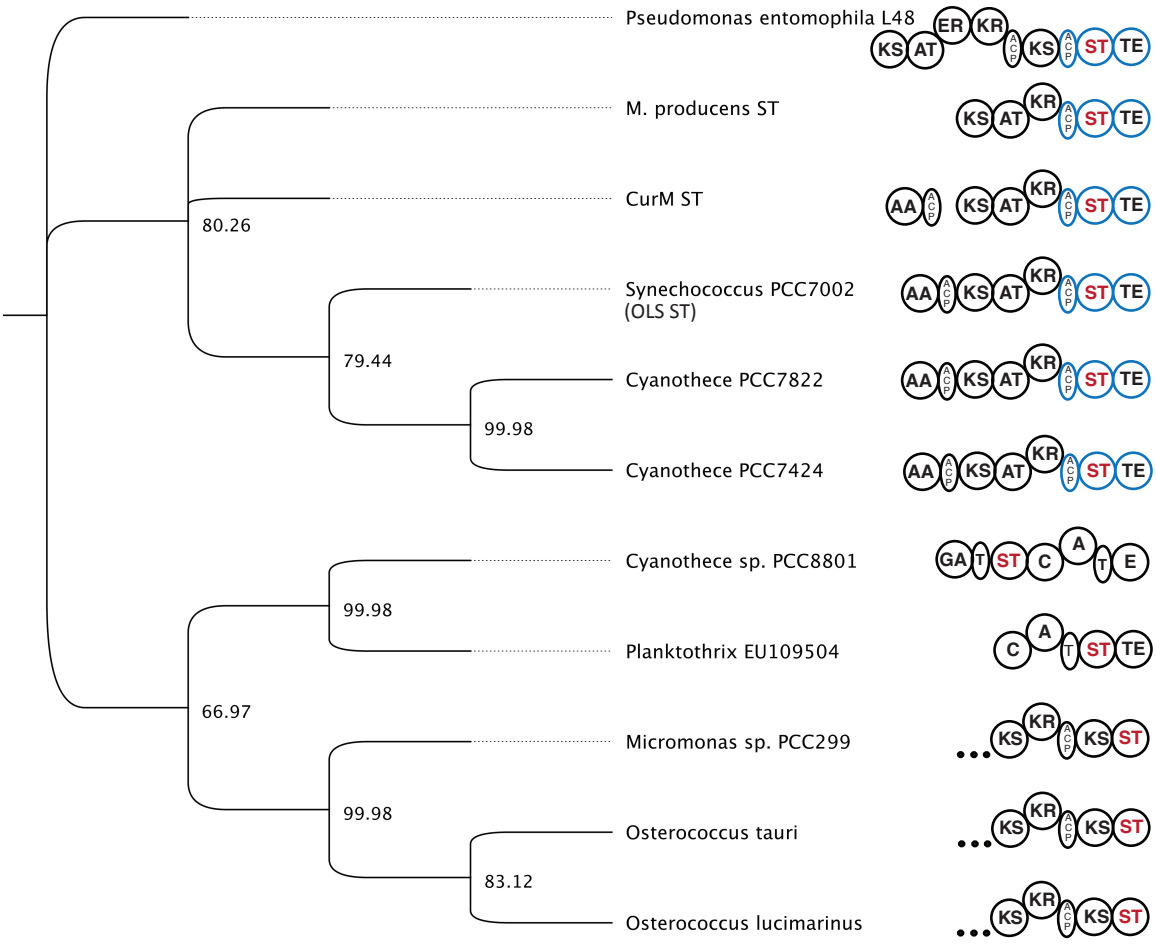


Figure 3.3 Phylogenetic analysis of activating STs and close homologs
ST domains are depicted in red and activating ACP-ST-TE tridomains are highlighted in blue. Domains neighboring the ST are depicted: ketosynthase (KS), acyl transferase (AT), ketoreductase (KR), acyl carrier protein (ACP), sulfotransferase (ST), thioesterase (TE), acyl activating (AA), glyceric acid loading (GA), thiolation (T), condensation (C), and adenylation (A). GenBank accession numbers for sequences not listed in Figure 3.2 are: *Planktothrix agardhii* NIVA-CYA 116 (ABI26077), *Cyanothece* PCC 8801 (YP_002372038), *Ostreococcus lucimarinus* CCE9901 (XP_001416378), *Micromonas* RCC299 (XP_002507643), *Ostreococcus tauri* COG3321 (XM_003074782). Phylogenetic analysis was performed by Dr. Lena Gerwick (Scripps Institution of Oceanography).

free fatty acid to ACP (53), consistent with the AA-ACP fusion. The central domains (ketosynthase (KS), acyltransferase (AT), ketoreductase (KR)), as in CurM, are presumed to catalyze a single extension of the fatty acyl chain using malonyl-CoA (AT, KS) with β-carbonyl group reduction by the KR to create a β-hydroxyacyl-ACP substrate for ST-TE mediated termination. This intriguing domain arrangement (AA-ACP-KS-AT-KR-ACP-ST-TE) (Figure 3.1A) suggests a pathway to convert fatty acids into terminal-alkene hydrocarbons with an odd number of carbon atoms due to the final decarboxylation step (27, 10). Indeed, the role of this biosynthetic system as an olefin synthase (OLS) was recently confirmed for the gene from *Synechococcus* PCC 7002 (52). Based on these findings it is evident that the unique PKS/FAS ST-TE offloading process has exciting potential applications in engineering pathways for hydrocarbon biofuel production and natural product structural diversification (100, 101).

Due to the surprising role of CurM PKS ST and OLS ST domains in hydroxy group activation for the terminal decarboxylative elimination reaction, we were motivated to obtain detailed structural information on these proteins. Previous crystal structures have primarily focused on mammalian STs, which share a common protein architecture, including a single catalytic domain, a common binding site for the sulfonate donor PAPS, and flexible loops surrounding the active site that become ordered upon PAP(S) binding (102). These structures and the PAP(S) binding mode led to the prevailing view of ST catalysis (103), in which a general base activates the acceptor substrate nucleophile for a concerted in-line transfer of sulfonate from PAPS.

Here, we report structure-function studies of activating STs from the curacin A pathway (CurM PKS ST domain) and from the *Synechococcus* PCC 7002 olefin synthase system (OLS ST domain, 47% sequence identity to CurM ST). Both activating STs accept substrates leading to long-chain hydrocarbon production, but the OLS ST excludes substrates bearing a C5-methoxy group (Figure 3.1B), which is present in the natural curacin chain elongation substrate. Crystal structures of both STs reveal the active-site residues, details of PAPS binding, and a dynamic active-site flap involved in substrate selectivity.

Methods

Cloning and site-directed mutagenesis.

A construct encoding CurM ST (CurM residues 1598–1917, here referred to as 1–320) was amplified from cosmid pLM17 (2) and inserted into the pMCSG7 vector encoding an N-terminal 6xHis tag. A construct encoding OLS ST (OLS residues 2121–2424, here referred to as 9–312) was amplified from synthetic DNA (GeneArt) and inserted into pMCSG7. Surface entropy reduction substitutions in CurM ST (E183A/E184A, K223A/K224A, Q259A/K260A, and E209A/E211A) were designed with assistance from the SERp server (104). Site directed mutagenesis of single amino acids was performed using the QuikChange protocol (Agilent). Flap chimeras were constructed using the QuikChange Lightening Multi Site Directed Mutagenesis Kit (Agilent). All constructs were verified by sequencing. CurM ACP was expressed as previously described (27).

Protein expression, and purification.

E. coli strain BL21(DE3) was transformed with expression plasmid, grown at 37˚C in 500 mL TB with 4% glycerol to an OD_{600} of 1.0, cooled to 20^{o}C, induced with IPTG (final concentration 0.2 mM), and grown for an additional 18 h. Selenomethionyl (SeMet) CurM $ST_{Q259A/K260A}$ was produced in the same *E. coli* strain in SelenoMet Medium (AthenaES) containing 100 μg mL^{-1} of seleno-DL-methionine.

All purification steps were performed at 4˚C. The cell pellet from 500 mL of cell culture was resuspended in 40 mL Buffer A (20 mM Tris pH 7.9, 500 mM NaCl, 10% glycerol) plus 20 mM imidazole and lysed by sonication, and the soluble fraction loaded onto a 5-mL HisTrap Ni NTA column (GE Healthcare). The proteins were eluted in Buffer A with a linear gradient of 20 to 650 mM imidazole. The 6xHis tag was removed by 2-h incubation with 1 mM DTT and tobacco etch virus (TEV) protease (1 mg protease per 50 mg ST) at room temperature. After overnight dialysis at 4^{o}C in Buffer A with 1 mM DTT, the remaining His-tagged proteins were removed by Ni-affinity chromatography, followed by size exclusion chromatography with a HiLoad 16/60 Superdex 200 column (GE Healthcare) pre-equilibrated with Buffer A. The purified proteins were concentrated to 15 mg/mL, flash frozen in liquid N_2, and stored at –80˚C. SeMet CurM $ST_{Q259A/K260A}$ was purified as the wild type with addition of 2 mM DDT to Buffer A during the size exclusion step. Of the 4 surface entropy reduction variants, 2 yielded enough soluble protein for crystallization trials (Q259A/K260A and

E209A/E211A). CurM ST variants used for assay were purified as above but using a single HisTrap Ni NTA purification step followed by overnight dialysis in Buffer A with yields similar to wild type. Yields per 500 mL culture were 20 mg for CurM ST, 60 mg for OLS ST, 12 mg for CurM $ST_{Q259A/K260A}$, and 4 mg for SeMet CurM $ST_{Q259A/K260A}$.

Crystallization.

Initial attempts to produce high-quality CurM ST crystals were not successful, so surface entropy reduction (SER) mutations (105) were made to promote crystallization. Double alanine substitutions were made at four sites of non-conserved, adjacent, polar amino acids; one double substitution (Q259A/K260A) resulted in crystals that diffracted to high resolution (1.6 Å). The activity of CurM $ST_{Q259A/K260A}$ was reduced about threefold relative to wild type (Table 3.1). Crystals of CurM $ST_{Q259A/K260A}$ grew at 4°C within 24–48 h by vapor diffusion in hanging drops from a 1:1 mix of protein stock (7 mg mL^{-1} ST in Buffer A, 2 mM PAP) and well solution (25–31% PEG MME 550, 2.5–12.5 mM $ZnSO_4$, 100 mM MES pH 6.5). Crystallization was dependent upon the presence of Zn^{2+} and PAP. SeMet protein crystallized in similar conditions to the native CurM $ST_{Q259A/K260A}$ with 1 mM DTT in the protein solution. OLS ST crystals grew at 20°C within 24 h from a 1:1 mixture of protein stock (11 mg mL^{-1} ST in Buffer A, 2 mM PAP) and well solution (27–30% PEG 1.5K, 100 mM MMT buffer (DL-malic acid, MES, Tris) pH 6.0). Crystals of OLS ST that diffracted beyond 4.0 Å were grown by microseeding. Crystals were harvested in loops and flash cooled in liquid N_2.

Table 3.1 Effect of Zn^{2+} and Gln259/Lys 260 Ala substitutions of CurM ST activity

	% WT
CurM ST WT	100.0[a]
100 μM Zn^{2+}	52.9
10 μM Zn^{2+}	74.3
1 μM Zn^{2+}	73.9
Q259A/K260A	34.6

[a]Raw HPLC chromatogram peak area for the 3-*R*-hydroxymyristoyl-ACP substrate and sulfated product were used to calculate the fraction of substrate sulfated and then normalized to a wild type zinc-free control.

Data collection and structure determination.

Data were collected at GM/CA beamline 23ID-D at the Advanced Photon Source (APS) at Argonne National Lab (Argonne, IL) (Table 3.2). A 2.1-Å single-wavelength anomalous diffraction dataset was collected at the wavelength of peak absorption at the selenium edge from a SeMet CurM $ST_{Q259A/K260A}$ crystal (Table 3.3). A 1.6-Å dataset was collected from a native CurM $ST_{Q259A/K260A}$ crystal at E = 12.000 keV (Table 3.4) and a 1.9-Å dataset was collected at E = 10.000 keV (Table 3.5) from a second crystal from the same drop. A 2.1-Å dataset was collected from an OLS ST crystal (Table 3.6). All data were processed using the HKL2000 suite (80). Using Phaser (81) in the PHENIX (82) software suite, anomalous scatterer sites were found for all eight Met residues from the one polypeptide chain in the asymmetric unit (average figure of merit (FOM) = 0.49), with two partially occupied Se sites for each of five Met side chains (Table 3.7). Five additional anomalous scatterers were found in locations that did not correspond to Met side chains. These five anomalous scatterers were identified as Zn^{2+} in an anomalous difference Fourier from a dataset collected at 10.000 keV (λ = 1.23984 Å), 341 eV above the Zn K absorption edge. After density modification in RESOLVE (83) (FOM = 0.77) a 90% complete model was built by AUTOBUILD (84) and completed manually in COOT (85). Waters were added automatically in COOT and manually edited. The 1.6-Å native dataset was used for refinement in REFMAC5 (86) from the CCP4 (87) suite using 5 TLS groups (88). Electron density was complete throughout the polypeptide chain for residues 26–308. Twenty-five residues at the N-terminus and 12 at the C-terminus were disordered (Figure 3.4). The OLS ST structure was solved by molecular replacement in Phaser (81) using CurM $ST_{Q259A/K260A}$ as a search model (47% sequence identity). OLS ST was refined using BUSTER (106) (Figure 3.5). Model quality was evaluated with MolProbity (107).

Table 3.2 Crystallographic summary

	CurM $ST_{Q259A/K260A}$ (SeMet)	CurM $ST_{Q259A/K260A}$ (native)	CurM $ST_{Q259A/K260A}$ (native)	OLS ST (native)
Diffraction Data				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P4_22_12$
X-ray source	APS 23-ID-B	APS 23-ID-B	APS 23-ID-B	APS 23-ID-D
a, b, c (Å)	46.0, 67.6, 118.5	45.8, 67.3, 118.0	45.9, 67.3, 118.5	131.4, 131.4, 47.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97948	1.03320	1.23984	0.97934
d_{min} (Å)	2.12 (2.20-2.12)[a]	1.62 (1.68-1.62)	1.95 (2.02-1.95)	2.15 (2.23-2.15)
Avg I/σ_I	26.5 (14.3)	17.5 (4.9)	17.0 (7.8)	24.1 (3.6)
R_{symm}[b]	0.098 (0.261)	0.074 (0.375)	0.085 (0.276)	0.063 (0.634)
Completeness	100 (99.9)	98.5 (89.6)	99.7 (99.4)	99.8 (100.0)
Avg. redundancy	14.5 (14.3)	6.6 (5.9)	7.2 (7.2)	8.7 (7.5)
Unique reflections	21,790	46,830	27,422	23,099
Refinement				
Data range (Å)		36.34-1.62		33.07 -2.15
No. reflections		46,772		22,991
R_{work}/R_{free}[c]		0.158/0.182		0.188/0.220
RMS deviations				
Bonds (Å)		0.007		.010
Angles (°)		1.130		1.06
Avg *B*-factors ($Å^2$)				
Protein		17.1		61.7
Water		33.5		60.7
Other		22.1		48.2
Ramachandran				
Allowed		100.0%		100.0%
Outliers		0.0%		0.0%
Protein Atoms		2325		2211
Water Molecules		365		96
Other Atoms		83		27

[a]Outermost shell in parentheses.
[b]Including anomalous differences.
[c]The R_{free} data set included a random 5% of reflections.

Table 3.3 Scaling statistics for SeMet CurM $ST_{Q259A/K260A}$

Shell Limit (Å)						
Lower	Upper	Average I	Avg. error	Chi^2	Linear R-factor[a]	Square R-factor
50.00	4.57	653.3	25.5	1.520	0.074	0.076
4.57	3.62	723.5	24.1	1.361	0.075	0.080
3.62	3.17	429.3	13.6	1.378	0.078	0.080
3.17	2.88	246.7	8.3	1.502	0.100	0.104
2.88	2.67	174.7	6.0	1.463	0.113	0.114
2.67	2.51	144.0	5.5	1.278	0.122	0.118
2.51	2.39	113.0	5.2	1.131	0.149	0.148
2.39	2.28	103.6	5.2	1.115	0.170	0.163
2.28	2.20	95.5	5.3	1.090	0.195	0.197
2.20	2.12	72.7	5.1	1.134	0.261	0.268
All	Reflections:	281.4	10.6	1.298	0.098	0.086

[a]Includes anomalous differences

Table 3.4 Scaling statistics for CurM $ST_{Q259A/K260A}$ collected at 12.000 keV

Shell Limit (Å)						
Lower	Upper	Average I	Avg. error	Chi^2	Linear R-factor	Square R-factor
50.00	3.49	1114.1	45.4	0.944	0.052	0.067
3.49	2.77	519.3	20.7	1.110	0.063	0.074
2.77	2.42	269.2	11.5	1.021	0.069	0.079
2.42	2.20	208.6	9.8	1.046	0.079	0.085
2.20	2.04	150.8	8.0	1.101	0.097	0.100
2.04	1.92	105.7	6.8	1.024	0.116	0.113
1.92	1.82	71.8	6.0	1.031	0.156	0.148
1.82	1.75	51.9	5.6	1.016	0.210	0.195
1.75	1.68	39.8	5.7	1.025	0.278	0.261
1.68	1.62	30.3	6.2	1.007	0.375	0.349
All	Reflections:	266.3	12.9	1.033	0.074	0.071

Table 3.5 Scaling statistics for CurM $ST_{Q259A/K260A}$ collected at 10.00 keV

Shell Limit (Å)						
Lower	Upper	Average I	Avg. error	Chi^2	Linear R-factor[a]	Square R-factor
50.00	4.20	4032.8	217.0	1.031	0.060	0.067
4.20	3.33	3630.3	184.5	1.014	0.068	0.075
3.33	2.91	1630.7	89.4	0.954	0.077	0.084
2.91	2.65	982.6	59.1	0.937	0.090	0.098
2.65	2.46	708.9	46.8	0.938	0.103	0.109
2.46	2.31	542.2	36.2	1.072	0.121	0.127
2.31	2.20	482.3	36.8	1.093	0.140	0.153
2.20	2.10	344.0	31.6	1.127	0.177	0.187
2.10	2.02	272.5	31.2	1.092	0.220	0.232
2.02	1.95	199.0	25.5	1.272	0.276	0.276
All	Reflections:	1313.6	77.4	1.052	0.085	0.077

[a]Includes anomalous differences

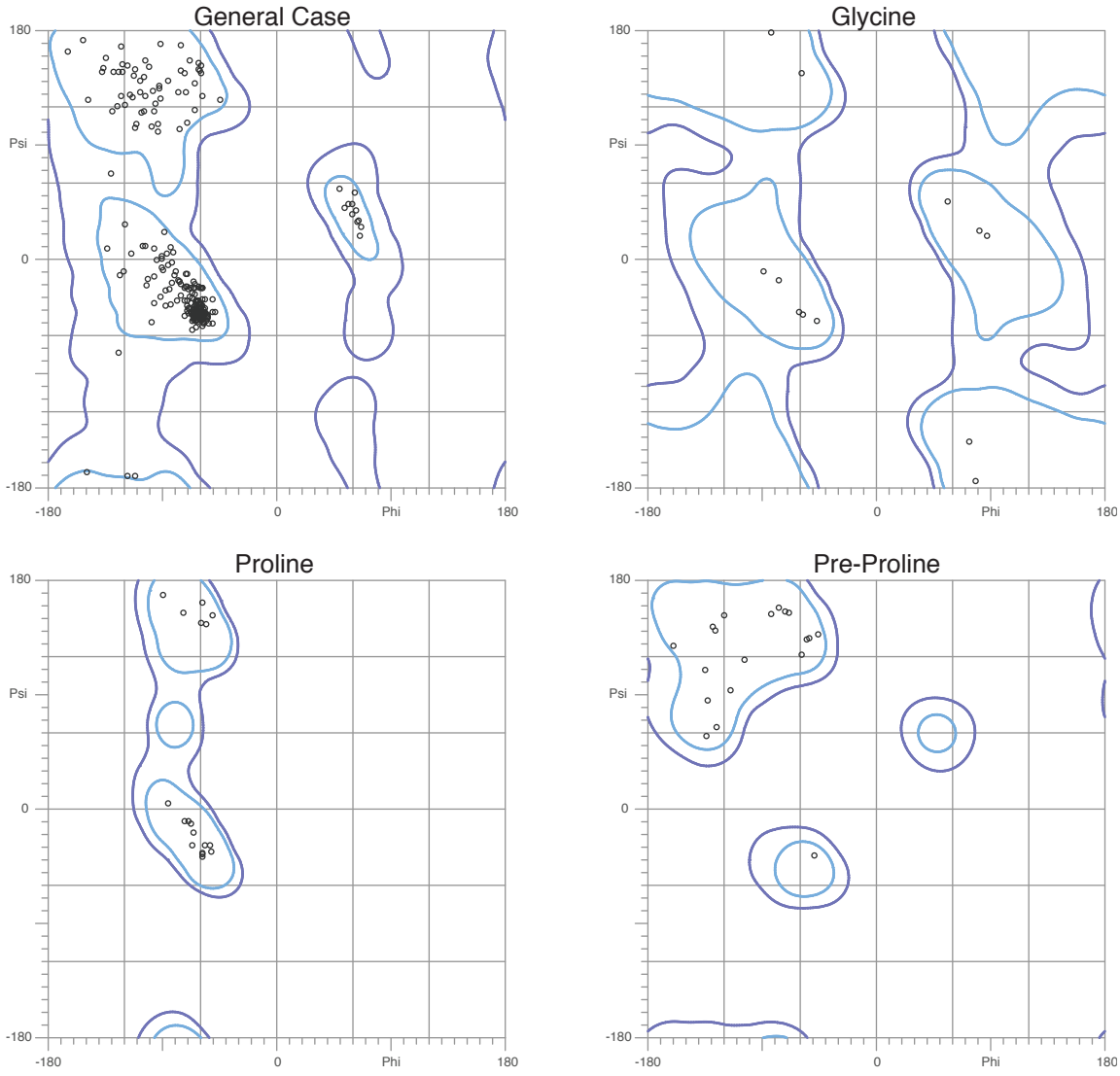
Table 3.6 Scaling statistics for OLS ST

Shell Limit (Å)						
Lower	Upper	Average I	Avg. error	Chi^2	Linear R-factor	Square R-factor
50.00	4.63	2545.1	85	0.797	0.035	0.039
4.63	3.68	1796.3	64.2	0.978	0.053	0.058
3.68	3.21	848.6	29.2	1.047	0.058	0.057
3.21	2.92	332.2	18.0	0.985	0.096	0.096
2.92	2.71	189.3	11.9	1.074	0.125	0.118
2.71	2.55	118.5	9.3	1.068	0.158	0.136
2.55	2.42	83.2	8.7	1.097	0.215	0.177
2.42	2.32	57.8	8.5	1.121	0.314	0.262
2.32	2.23	42.9	8.7	1.218	0.452	0.359
2.23	2.15	32.9	9.1	1.247	0.633	0.505
All	Reflections:	626	25.9	1.056	0.063	0.051

Table 3.7 Phasing statistics for CurM ST

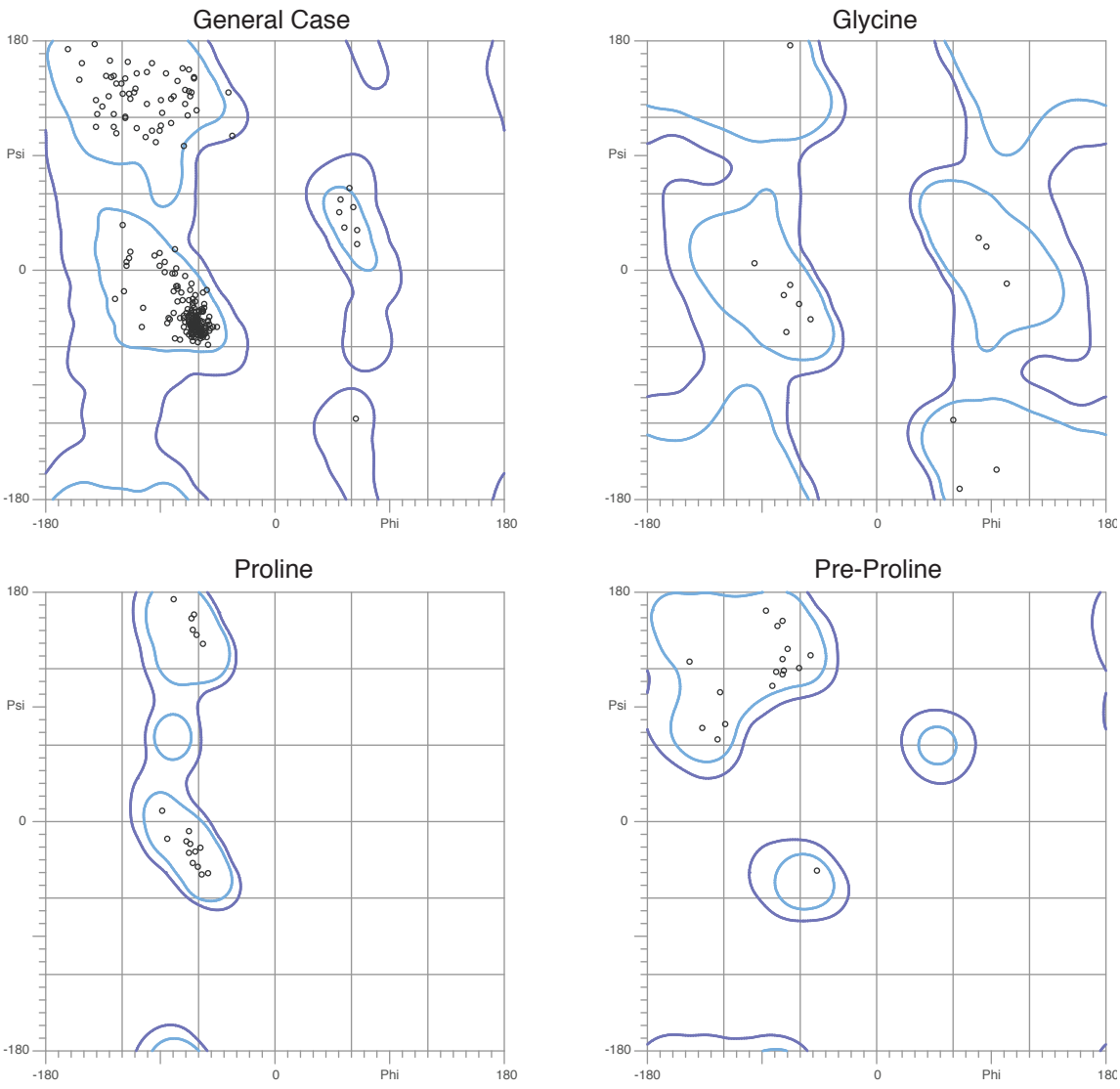
Anomalous signal	0.074
Sites found	18
Bayes-CC[a]	61.5
FOM[b]	0.49
LLG[c]	-224,532

[a]Bayesian estimate of the quality of experimental electron density maps.
[b]Final figure of merit of phasing
[c]Log-likelihood gain



98.6% (287/291) of all residues were in favored (98%) regions.
100.0% (291/291) of all residues were in allowed (>99.8%) regions.

Figure 3.4 Ramachandran analysis of CurM ST
Ramachandran plots of final refined model of CurM ST (PDB code = 4GBM). Plots were generated using MolProbity (http://kinemage.biochem.duke.edu) (89).



98.1% (264/269) of all residues were in favored (98%) regions.
100.0% (269/269) of all residues were in allowed (>99.8%) regions.

Figure 3.5 Ramachandran analysis of OLS ST
Ramachandran plots of final refined model of OLS ST (PDB code = 4GOX). Plots were generated using MolProbity (http://kinemage.biochem.duke.edu) (89).

Crystals of both STs contained one polypeptide in the asymmetric unit. OLS ST amino acid numbering within the ST domain was assigned to match with CurM ST numbering, as there are no gaps in the sequence and structure alignments. Electron density is continuous for CurM ST residues 26 to 308, and OLS ST termini are ordered in a similar range, 18 to 308. The OLS ST crystal structure has two disordered internal loops (177–184 and 255–263), which are ordered in CurM ST. Zn^{2+} was essential to

CurM ST crystal formation. A total of five Zn positions were identified by Zn anomalous scattering (Figure 3.6A). Weak anomalous scattering peaks at other positions in the tetrahedral coordination sphere of Zn^{2+} (Table 3.6) were interpreted as Cl^- ligands.

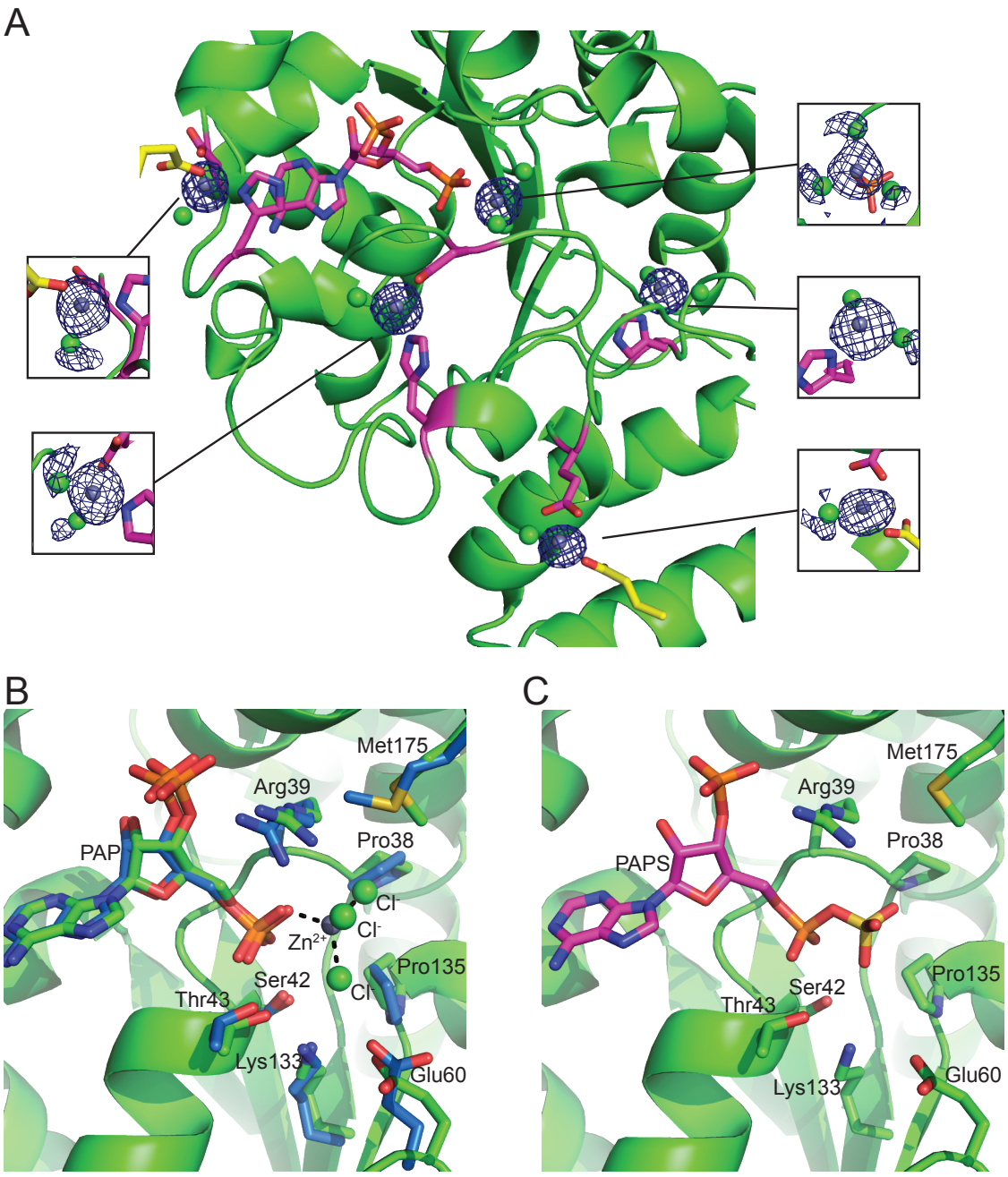


Figure 3.6 Zinc binding in CurM ST crystals
A. An anomalous difference Fourier map contoured at 5σ from a dataset collected at 10.000 keV (1.23984 Å), 341 eV above the Zn K absorption edge. Peaks for the five zinc bound to CurM ST are shown. Amino acid ligands are shown in stick (magenta C, and yellow C for residues from a neighboring molecule in the crystal lattice). Small panels show anomalous difference density contoured at 2.5σ. Peaks adjacent to the zinc sites were interpreted as chloride ligands (green). **B.** Zinc bound to PAP in CurM ST superimposed with OLS ST. Zn^{2+} (gray) coordinates three Cl^- ligands (green) and one PAP 5'-phosphate oxygen (green C, red O, blue N, orange P). Residues surrounding the $ZnCl_3^-$ are shown in stick for CurM ST (green C) and OLS ST (blue C). **C.** PAPS in the CurM ST active site. View as in B with modeled PAPS (magenta C), using PAP-$ZnCl_3^-$ as a guide.

Sequence alignment, structure alignment, and substrate modeling.

The search for ST homologs was done with BLAST (90). The DALI server (108) was used to identify the closest known structures to activating STs. MUSCLE (91) was used for multiple sequence alignment. PyMOL was used to align structures and prepare figures (109). The PRODRG2 server (93) was used to generate starting models and topology files for the substrates, which were docked manually using the position of the 3-OST substrate (1T8U (110)) as a guide. Sequences were aligned for phylogenetic analysis using the MAAFT alignment algorithm followed by tree building in MrBayes (111) using the Whelan-Goldman model and gamma distribution variation. The most appropriate model was determined by adding the alignment to the TOPALi modeling program (http://www.topali.org/). Phylogenetic analysis was performed by Dr. Lena Gerwick (Scripps Institution of Oceanography).

Enzyme assay.

CurM ST and OLS ST were assayed using a modification of our previous protocols (27, 99). The synthesis of (3*R*)-3-hydroxy-5-methoxymyristoyl-CoA and (3*S*)-3-hydroxy-5-methoxymyristoyl-CoA was previously described (27). (*R*)-3-hydroxymyristoyl-CoA and (*R*)-3-hydroxydodecanoyl-CoA were synthesized as described in the following Chemical Synthesis section. Apo-CurM ACP was loaded with a substrate analog by 2-h incubation of 50 μM apo-ACP with 100 μM CoA-linked substrate ((3*R*)-3-hydroxy-5-methoxymyristoyl-CoA (27), (3*S*)-3-hydroxy-5-methoxymyristoyl-CoA (27), (*R*)-3-hydroxymyristoyl-CoA, and (*R*)-3-

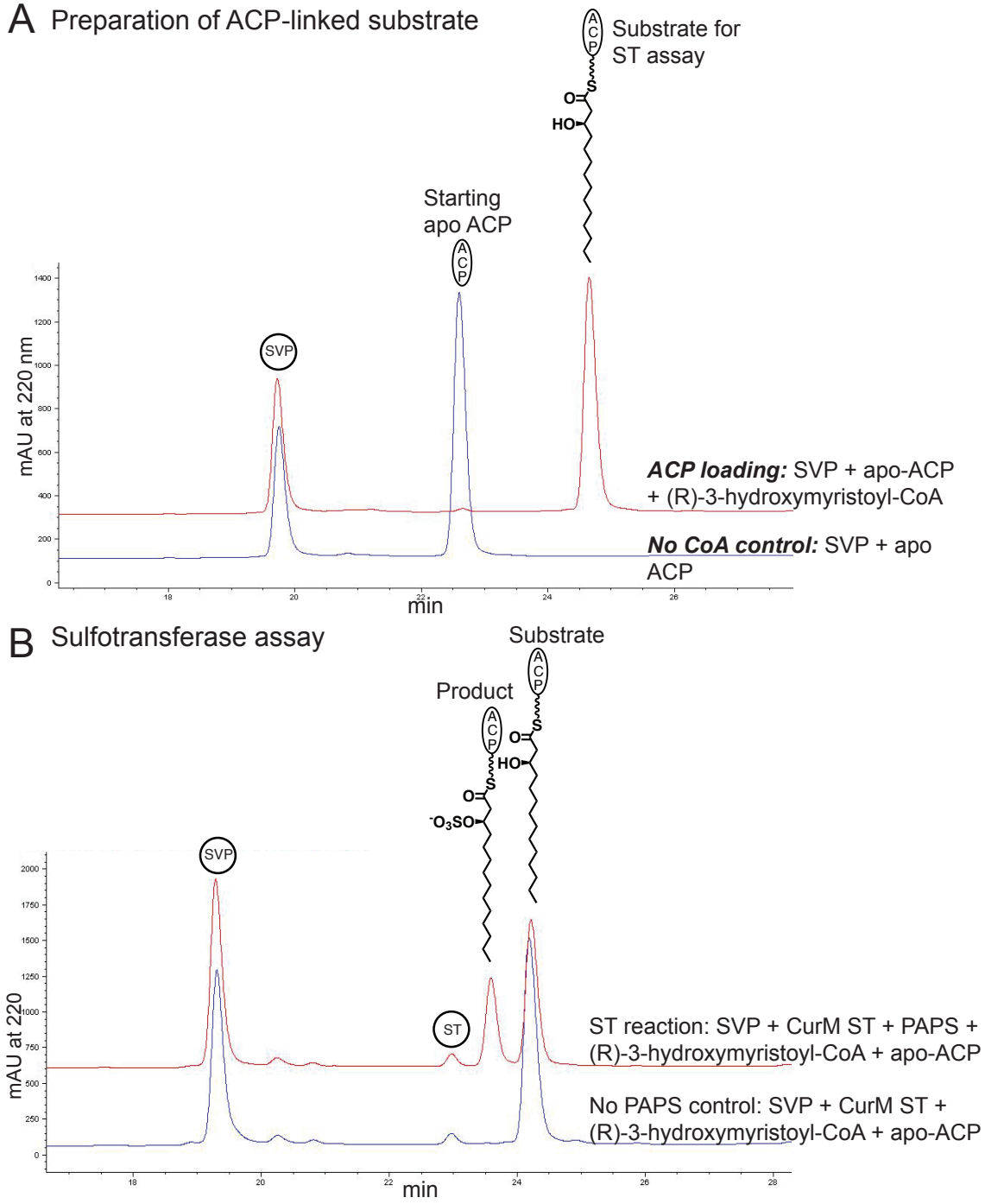
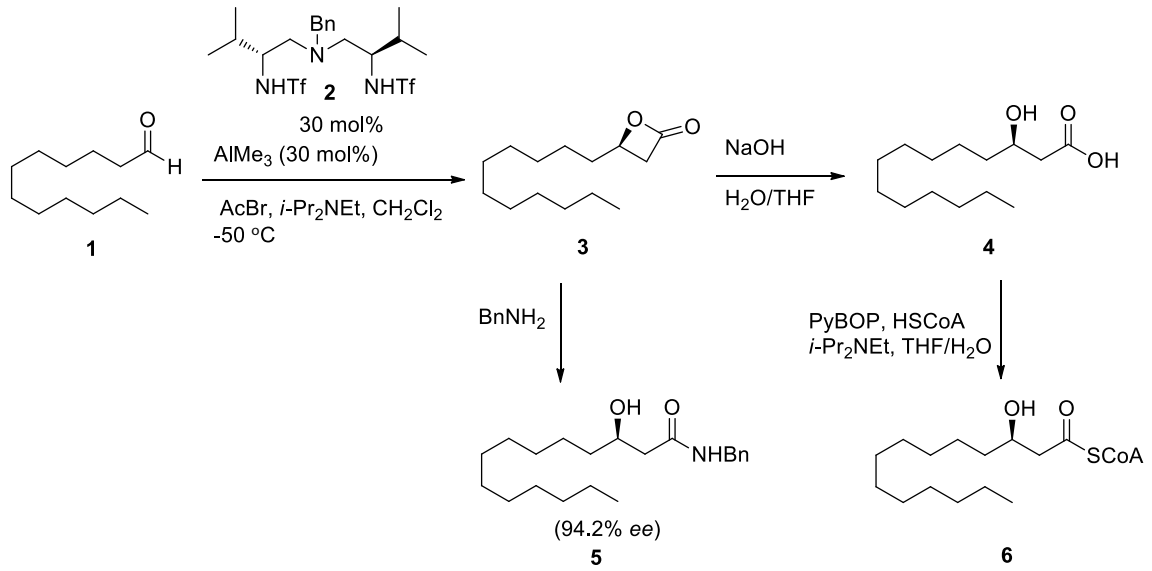

A Preparation of ACP-linked substrate
Substrate for ST assay
Starting apo ACP
SVP
mAU at 220 nm
min
ACP loading: SVP + apo-ACP + (R)-3-hydroxymyristoyl-CoA
No CoA control: SVP + apo ACP
B Sulfotransferase assay
Substrate
Product
SVP
ST
mAU at 220
min
ST reaction: SVP + CurM ST + PAPS + (R)-3-hydroxymyristoyl-CoA + apo-ACP
No PAPS control: SVP + CurM ST + (R)-3-hydroxymyristoyl-CoA + apo-ACP

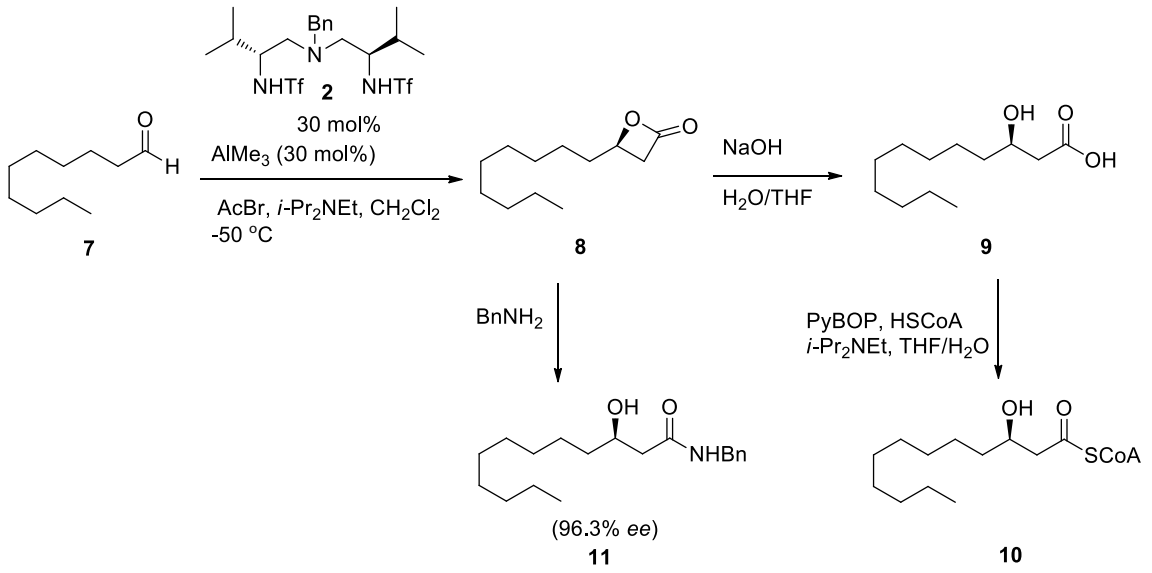
Figure 3.7 HPLC detection of ST substrates and products
A. The ACP loading reaction analyzed by reverse-phase HPLC. Svp was combined with apo-ACP and (*R*)-3-hydroxymyristoyl-CoA to load the ACP with (*R*)-3-hydroxymyristoyl-phosphopantethiene. The lower trace is a CoA-free control. **B.** The ST reaction analyzed by HPLC. The one-pot assay mixture was separated by reverse-phase HPLC resulting in peaks for Svp, (*R*)-3-hydroxymyristoyl-ACP, and the product, (*R*)-3-sulfomyristoyl-ACP.

hydroxydodecanoyl-CoA), 10 μM *Streptomyces verticillus* Svp (94), 10 mM $MgCl_2$, and 100 mM Tris pH 7.9 at 30°C. Complete loading was confirmed 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 30% to 90% CH_3CN (0.1% CF_3CO_2H)/H_2O (0.1% CF_3CO_2H) over 45 min (Figure 3.7A). After exchange into Buffer A and concentration (Amicon Ultra 10 kDa concentrators, Millipore), substrate-loaded ACP was flash frozen and stored at –80°C. It was not possible to reach saturating concentrations of ACP-linked substrates, so standard assay conditions were developed where nearly complete sulfonation of substrates was achieved with the wild type CurM ST. Reaction times were adjusted for the slower OLS ST. The reaction was initiated by the addition of ST (1 μM) into 100 μM loaded ACP, 1.75 mM PAPS (Sigma), 100 mM Tris pH 7.9 at room temperature. After 5 min (CurM ST) or 3 h (OLS ST), the reactions were quenched with 10% formic acid. Conversion of β-hydroxy-ACP to sulfated-ACP was quantified by HPLC detection of ACP species as described above (Figure 3.7B). Results were quantitated as the sulfated-ACP peak area as a fraction of the total ACP peak area (substrate + product). Under the conditions tested, CurM ST is more active than OLS ST, with 75% sulfonation of the ACP-bound substrate in 5 minutes whereas OLS ST converts approximately 50% in 3 hours.

Chemical synthesis



Scheme 1. Synthesis of 3-(*R*)-tetradecanoylCoA (**6**) and amide **5**

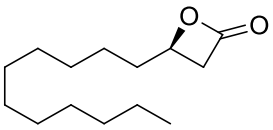


Scheme 1. Synthesis of 3-(*R*)-dodecanoylCoA (**10**) and amide **11**

Chemical synthesis – general

The chemical synthesis was performed by Dr. Sarang Kulkarni in the lab of Dr. Peter Wipf (University of Pittsburgh). All reactions were performed under an N_2 atmosphere and all glassware was flame dried prior to use. CH_2Cl_2 and toluene were dried by passing through a column of activated alumina and degassed prior to use by freeze-pump-thaw method. Reactions carried out at -78°C employed a CO_2/acetone bath. THF and Et_2O were distilled over sodium/benzophenone ketyl.

Reactions were monitored by TLC analysis (pre-coated silica gel 60 F254 plates, 250 μm layer thickness) and visualization was accomplished with a 254 nm UV light and by staining with a PMA solution (5 g of phosphomolybdic acid in 100 mL of 95% EtOH), *p*-anisaldehyde solution (2.5 mL of *p*-anisaldehyde, 2 mL of AcOH, and 3.5 mL of conc. H_2SO_4 in 100 mL of 95% EtOH), Vaughn's reagent (4.8 g of $(NH_4)_6Mo_7O_{24}•4H_2O$ and 0.2 g of $Ce(SO_4)_2$ in 100 mL of a 3.5 N H_2SO_4 solution) or a $KMnO_4$ solution (1.5 g of $KMnO_4$ and 1.5 g of K_2CO_3 in 100 mL of a 0.1% NaOH solution). Flash chromatography on SiO_2 was used to purify the crude reaction mixtures. ^1H NMR spectra were obtained at 300 or 400 MHz in $CDCl_3$ unless otherwise noted. Chemical shifts were reported in parts per million with the residual solvent peak used as an internal standard. ^1H NMR spectra were obtained and are tabulated as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). ^{13}C NMR spectra were obtained at 76 or 100 MHz using a proton-decoupled pulse sequence with a d1 of 3 sec, and are tabulated by observed peak. IR spectra were obtained on an IdentifyIR-ATR spectrometer.

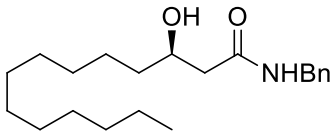


3

*Chemical synthesis, (R)-4-Undecyloxetan-2-one (**3**)*

To a flame-dried, N_2-cooled flask equipped with a stir bar was added a solution of the triamine ligand **2** (112) (0.371 g, 0.684 mmol, 30 mol%) in CH_2Cl_2 (12 mL). The mixture was stirred at room temperature and treated slowly with a solution of $AlMe_3$ (49.0 mg, 0.684 mmol, 30 mol%) in CH_2Cl_2 (3 mL). The reaction mixture was stirred at

room temperature for 2 h, cooled to -50°C and treated with *i*-Pr_2NEt (0.680 mL, 3.87 mmol), freshly distilled acetyl bromide (0.320 mL, 4.33 mmol) and dodecanal (0.550 mL, 2.28 mmol). The resultant pale yellow solution was stirred for 14 h at -45°C, warmed to room temperature, and poured into a separatory funnel containing 0.1 N HCl (25 mL). The organic layer was successively washed with sat. $NaHCO_3$ and brine. The organic phase was dried (Na_2SO_4) and the solvent was removed *in vacuo* to afford a yellow oil that was purified by chromatography on SiO_2 (hexanes:EtOAc, 9:1) to provide **3** (the enantiomeric excess of lactone **3** was determined by chiral HPLC analysis of amide **5**, obtained by ring opening of **3** with benzyl amine) (0.351 g, 68% yield) as a colorless oil: $[\alpha]_D$ +15.0 (*c* 1.55, $CHCl_3$); IR (ATR) 2921, 2851, 1825, 1446, 1144, 1123 cm^{-1}; ^1H NMR ($CDCl_3$, 300 MHz) δ 4.54-4.46 (m, 1 H), 3.50 (dd, 1 H, *J* = 5.7, 16.2 Hz), 3.05 (dd, 1 H, *J* = 4.5, 16.2 Hz), 1.89-1.68 (m, 2 H), 1.26 (app s, 20 H), 0.87 (t, 3 H, *J* = 6.9 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 168.6, 71.4, 42.9, 34.7, 32.0, 29.7, 29.6, 29.5, 29.4, 29.2, 25.0, 22.8, 14.2.

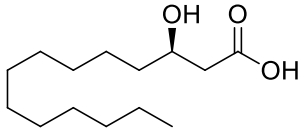


5

Chemical synthesis, (R)-N-Benzyl-3-hydroxytetradecanamide (**5**)

A mixture of lactone **3** (53 mg, 0.23 mmol) and benzylamine (0.13 mL, 1.2 mmol) was stirred at room temperature for 16 h. The reaction mixture was diluted with EtOAc (10 mL) and washed with 1 N HCl (2 x 10 mL), water (10 mL) and brine (10 mL). The organic layer was dried (Na_2SO_4) and concentrated under reduced pressure to give a brown residue, which was purified by chromatography on SiO_2 (hexanes:EtOAc, 1:1) to give **5** (59 mg, 76%) as a white solid. The enantiomeric excess of the **5** was determined to be 94.2% *ee* by chiral HPLC (ChiralCel OD, 4.6 mm X 250 mm, hexane/isopropanol (90/10, 30 min), 1 mL/min, λ = 221 nm, $t_R$major = 9.71 min, t_R minor = 12.33 min): Mp 90.4-92.6 °C (CH_2Cl_2); $[\alpha]_D$ -14.9 (*c* 1.06, $CHCl_3$); IR (ATR) 3295, 2924, 2850, 1644, 1567, 1424 cm^{-1}; ^1H NMR ($CDCl_3$, 400 MHz) δ 7.33-7.24 (m, 5 H), 6.45 (s, 1 H), 4.41 (d, 2 H, *J* = 5.7 Hz), 3.97 (br s, 1 H), 3.69 (br s, 1 H), 2.37 (dd, 1 H, *J* = 2.6, 15.3 Hz), 2.26 (dd, 1 H, *J* = 9.0, 15.2 Hz), 1.53-1.35 (m, 3 H), 1.25 (app s, 17 H), 0.88 (t, 3 H, *J* = 6.6

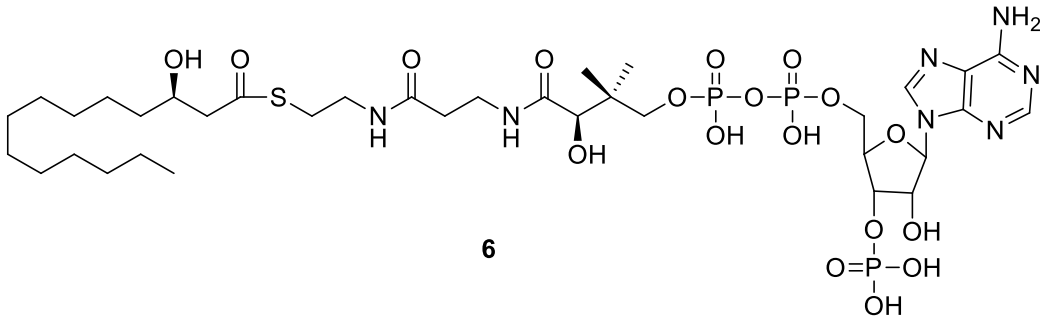
Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 172.5, 138.1, 128.8, 127.8, 127.6, 68.8, 43.5, 42.5, 37.1, 32.0, 29.7, 29.7, 29.7, 29.7, 29.6, 29.4, 22.8, 14.2; HRMS (TOF MS ES+) *m/z* calcd for $C_{21}H_{35}NO_2Na$ (M+Na) 356.2564, found 356.2594.



4

Chemical synthesis, (R)-3-Hydroxytetradecanoic acid (**4**)

To a stirred solution of **2** (112) (200 mg, 0.884 mmol) in THF (1.5 mL) at room temperature was added a solution of NaOH (117 mg, 2.92 mmol) in H_2O (1 mL). The reaction mixture was stirred at room temperature for 2 h, diluted with H_2O (5 mL) and washed with Et_2O (20 mL) to remove organic impurities. The ether layer was discarded and the aqueous layer was acidified with 2 N HCl (3 mL). The resulting suspension was extracted with Et_2O (2 x 15 mL). The combined organic layers were dried (Na_2SO_4), and the solvent was removed under reduced pressure to give **4** (188 mg, 87%) as a white solid: Mp 71.3-73.2 °C (CH_2Cl_2); $[\alpha]_D$ -14.4 (*c* 1.15, $CHCl_3$); IR (ATR) 3560, 2918, 2844, 1681, 1466, 1293, 1224 cm^{-1}; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 4.05-4.00 (m, 1 H), 2.57 (dd, 1 H, *J* = 3.1, 16.6 Hz), 2.47 (dd, 1 H, *J* = 9.0, 16.6 Hz), 1.58-1.43 (m, 3 H), 1.27 (app s, 17 H), 0.88 (t, 3 H, *J* = 6.6 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 178.1, 68.2, 41.2, 36.6, 32.0, 29.8, 29.7, 29.7, 29.7, 29.6, 29.5, 25.6, 22.8, 14.2; HRMS (TOF MS ES-) *m/z* calcd for $C_{28}H_{55}O_6$ (2M-H) 487.3999, found 487.4014.

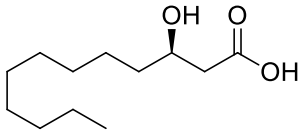


6

Chemical synthesis, (R)-3-Hydroxytetradecanoyl-CoA (**6**)

To a solution of acid **4** (1.59 mg, 0.00651 mmol) in THF (0.2 mL) was added a solution of PyBOP (13.5 mg, 0.0260 mmol) and Hünig's base (9.00 μL, 0.0521 mmol). To this mixture was added a solution of Coenzyme A free acid trihydrate (10.0 mg,

0.0131 mmol) in H_2O (0.2 mL). The resulting solution was stirred at room temperature for 5 h, lyophilized, and the resulting white solid was washed with THF to remove organic impurities. The residue was purified by reverse phase HPLC (Dynamax column (10 X 250 mm) with a gradient from 60% MeOH/40% H_2O (10 mM NH_4OAc) to 90% MeOH/10% H_2O (10 mM NH_4OAc), over 20 min and hold for 10 min, at a flow rate of 1 mL/min, t_R = 21.45 min). The solvent was evaporated under reduced pressure. The residue was dissolved in water and lyophilized to give **6** (3.8 mg, 59%) as a white solid: ^{1}H NMR (D_2O, 400 MHz) δ 8.53 (s, 1 H), 8.23 (s, 1 H), 6.13 (d, 1 H, *J* = 6.8 Hz), 4.54 (br s, 1 H), 4.23-4.19 (m, 2 H), 4.09-3.98 (m, 2 H), 3.79 (dd, 1 H, *J* = 4.8, 9.8 Hz), 3.51 (dd, 1 H, *J* = 4.7, 9.7 Hz), 3.41 (t, 2 H, *J* = 6.6 Hz), 3.31 (t, 2 H, *J* = 6.0 Hz), 3.04-2.93 (m, 2 H), 2.79-2.67 (m, 2 H), 2.39 (t, 2 H, *J* = 6.5 Hz), 1.44-1.40 (m, 2 H), 1.24-1.18 (m, 19 H), 0.84 (s, 3 H), 0.81 (t, 3 H, *J* = 6.7 Hz), 0.70 (s, 3 H); ^{13}C NMR (CD_3OD, 150 MHz) δ 200.5, 175.6, 174.3, 156.3, 155.1, 153.5, 152.2, 150.1, 140.9, 87.8, 84.3, 75.0, 72.9, 69.3, 66.3, 55.2, 51.2, 43.5, 39.7, 39.4, 37.9, 36.5, 36.3, 32.3, 30.8, 30.7 (2C), 30.4, 29.1, 26.5, 23.6, 22.0, 19.1, 18.7, 17.3, 14.9, 13.1; HRMS (TOF MS ES+) *m/z* calcd for $C_{35}H_{62}N_7O_{18}P_3SK$ (M+K) 1032.2722, found 1032.2648.



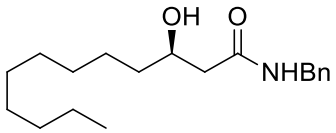
9

Chemical synthesis, (R)-3-Hydroxydodecanoic acid (**9**)

To a flame-dried, N_2-cooled flask equipped with a stirbar was added a solution of triamine ligand **2** (112) (475 mg, 0.877 mmol, 30 mol%) in CH_2Cl_2 (12 mL), followed slowly by a solution of $AlMe_3$ (63.0 mg, 0.877 mmol, 30 mol%) in CH_2Cl_2 (5 mL). The resulting solution was stirred at room temperature for 2 h, cooled to -50°C and treated with Hünig's base (0.870 mL, 4.97 mmol), freshly distilled acetyl bromide (0.410 mL, 5.56 mmol) and a solution of decanal (457 mg, 2.92 mmol) in CH_2Cl_2 (5 mL). The pale yellow reaction mixture was stirred for 14 h at -50°C, warmed to room temperature, and poured into a separatory funnel containing 0.1 N HCl (25 mL). The organic layer was washed with sat. $NaHCO_3$ and brine. The organic phase was dried (Na_2SO_4) and the solvent was removed *in vacuo* to afford a yellow oil, which was used in the following

step without further purification. (The enantiomeric excess of lactone **8** was determined by chiral HPLC analysis of amide **11**, obtained by ring opening of **8** with benzyl amine.)

To a stirred solution of above crude material in THF (6 mL) at room temperature was added a solution of NaOH (385 mg, 9.63 mmol) in H_2O (3 mL). The reaction mixture was stirred at room temperature for 2 h, diluted with H_2O (10 mL) and washed with Et_2O (3 x 20 mL) to remove organic impurities. The ether layer was discarded and to the aqueous layer was added 2 N HCl (10 mL). The resulting suspension was extracted with Et_2O (3 x 30 mL). The combined organic layers were dried (Na_2SO_4), and the solvent was removed under reduced pressure to give **9** (413 mg, 65% yield, 2 steps) as a white solid: Mp 59.8-62.0 °C (CH_2Cl_2); $[\alpha]_D$ -15.9 (*c* 1.00, $CHCl_3$); IR (ATR) 3560, 2919, 2844, 1676, 1466, 1293 cm^{-1}; ^1H NMR ($CDCl_3$, 400 MHz) δ 4.05-4.01 (m, 1 H), 2.56 (dd, 1 H, *J* = 3.1, 16.5 Hz), 2.46 (dd, 1 H, *J* = 9.0, 16.5 Hz), 1.58-1.43 (m, 3 H), 1.26 (app s, 13 H), 0.87 (t, 3 H, *J* = 6.6 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 178.1, 68.2, 41.2, 36.6, 32.0, 29.7, 29.7, 29.6, 29.4, 25.6, 22.8, 14.2.

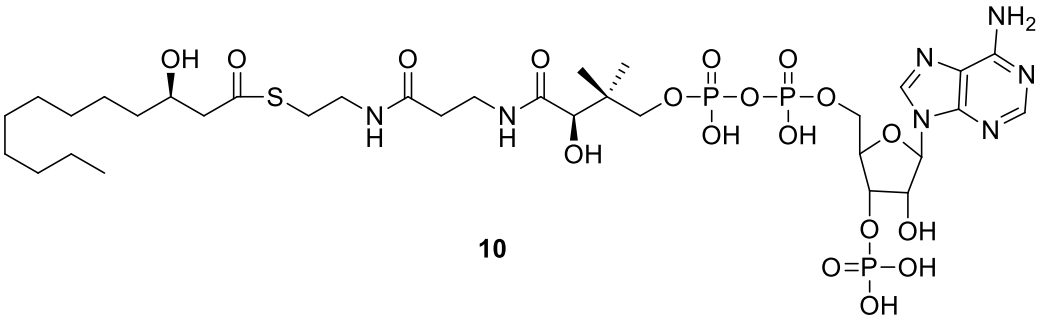


11

Chemical synthesis, (R)-N-Benzyl-3-hydroxydodecanamide (**11**)

A mixture of crude lactone **8** (50 mg) and benzylamine (0.140 mL, 1.26 mmol) was stirred at room temperature for 16 h. The reaction mixture was diluted with EtOAc (10 mL) and washed with 1 N HCl (2 x 10 mL), water (10 mL) and brine (10 mL). The organic layer was dried (Na_2SO_4) and concentrated under reduced pressure to give a brown residue, which was purified by chromatography on SiO_2 (hexanes:EtOAc, 1:1) to give **11** (42 mg) as a white solid. The enantiomeric excess of **11** was determined to be 96.3% *ee* by chiral HPLC (ChiralCel OD, 4.6 mm X 250 mm, hexane/isopropanol (90/10, 30 min), 1 mL/min, λ = 254 nm, $t_R$major = 9.97 min, t_R minor = 12.99 min): Mp 83.4-84.3 °C (hexanes:EtOAc, 1:1); $[\alpha]_D$ -15.8 (*c* 1.04, $CHCl_3$); IR (ATR) 3920, 2919, 2850, 1644, 1560, 1446 cm^{-1}; ^1H NMR ($CDCl_3$, 400 MHz) δ 7.33-7.24 (m, 5 H), 6.44-6.40 (m, 1 H), 4.41 (d, 2 H, *J* = 5.7 Hz), 4.00-3.95 (m, 1 H), 3.67 (br s, 1 H), 2.37 (dd, 1 H, *J* = 2.6, 15.3 Hz), 2.26 (dd, 1 H, *J* = 9.0, 15.2 Hz), 1.53-1.35 (m, 3 H), 1.25 (app s, 14 H), 0.88 (t,

3 H, J = 6.6 Hz); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 172.5, 138.1, 128.8, 127.8, 127.6, 68.8, 43.5, 42.6, 37.0, 32.0, 29.7, 29.6, 29.4, 25.6, 22.8, 14.2.



10

Chemical synthesis, (R)-3-Hydroxydodecanoyl-CoA (**10**)

To a stirred solution of acid **9** (1.89 mg, 0.00876 mmol) in THF (0.2 mL) was added PyBOP (18.2 mg, 0.0350 mmol) and Hünig's base (13.0 μL, 0.0701 mmol), followed by a solution of coenzyme A free acid trihydrate (14.4 mg, 0.0175 mmol) in H_2O (0.2 mL). The reaction mixture was stirred at room temperature for 5 h, lyophilized, and the resulting white solid was washed with THF to remove organic impurities. The white residue was purified by RP-HPLC (Dynamax C-18 (10 X 250 mm), gradient from 60% MeOH/40% H_2O (10 mM NH_4OAc) to 90% MeOH/10% H_2O (10 mM NH_4OAc) over 20 min and held for 10 min, at a flow rate of 1 mL/min, t_R = 20.18 min). The solvent was evaporated under reduced pressure. The residue was dissolved in water and lyophilized to give **10** (4.2 mg, 50%) as a white solid: ^{1}H NMR (D_2O, 400 MHz) δ 8.46 (s, 1 H), 8.12 (s, 1 H), 6.07 (d, 1 H, J = 6.0 Hz), 4.49 (br s, 1 H), 4.14 (br s, 2 H), 3.97-3.92 (m, 2 H), 3.74 (dd, 1 H, J = 4.8, 9.6 Hz), 3.55 (dd, 1 H, J = 4.8, 9.6 Hz), 3.34 (t, 2 H, J = 5.2 Hz), 3.24 (t, 2 H, J = 6.0 Hz), 2.93-2.89 (m, 2 H), 2.70-2.59 (m, 2 H), 2.32 (t, 2 H, J = 6.8 Hz), 1.35-1.32 (m, 2 H), 1.23-1.09 (m, 17 H), 0.79 (s, 3 H), 0.76 (t, 3 H, J = 6.4 Hz), 0.66 (s, 3 H); ^{13}C NMR (CD_3OD, 150 MHz) δ 201.3, 175.6, 174.6, 154.9, 151.7, 141.6, 87.8, 84.3, 82.1, 78.4, 75.0 (2C), 72.9, 69.4, 66.2, 51.8, 39.6, 39.3, 37.5, 36.4, 36.3, 32.6, 30.2, 30.1, 30.0, 29.1, 26.1, 23.4, 22.0, 19.1, 14.8.

Accession numbers.

Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4GBM for CurM ST and 4GOX for OLS ST).

Results

Activating STs are prototypes of a family.

In addition to the eight activating STs previously identified within ACP-ST-TE tridomains (10), at least five additional ST sequences of high identity to CurM ST (31%-44%) are present in sequence databases (Figure 3.3). These STs occur in the context of microbial natural product biosynthetic pathways, but unlike CurM ST, they lack a downstream decarboxylating TE. Rather, the genomic context suggests that these STs produce sulfated natural products, as has been detected for *Planktothrix* NIVA CYA 116 (41, 42). As we were most interested in STs with an activating functionality, the ST from the olefin synthase (OLS) in *Synechococcus* PCC 7002 (52) in addition to CurM ST from the curacin A PKS pathway were selected for further study.

The CurM PKS ST was demonstrated to be an activating ST (27) but no OLS ST has been characterized biochemically. Recombinant excised domains for CurM ST and OLS ST were purified as soluble, monomeric proteins and sulfotransferase activity was assayed (27). Briefly, β-hydroxyacyl substrates were loaded onto CurM ACP, and following reaction with ST, ACP products were analyzed by HPLC (Figure 3.1B, Figure 3.7). OLS ST reacted with two OLS substrates that would produce hydrocarbons when offloaded ((*R*)-3-hydroxymyristoyl-ACP and (*R*)-3-hydroxydodecanoyl-ACP) (Table 3.8), as expected from previous studies (52). However to our surprise, the OLS ST had no detectable activity toward either of the curacin PKS ST substrate analogs that bear a C5-methoxy substituent ((3*R*/*S*,5*R,S*)-3-hydroxy-5-methoxymyristoyl-ACP) (Table 3.8) under the assay conditions. In contrast, CurM ST acted similarly upon the curacin PKS ST substrate mimics and the OLS substrates but with a threefold preference for the natural *R*-β-hydroxy enantiomer (Table 3.8). OLS ST displayed a slight preference for the OLS substrate with the longer (14-carbon) alkyl chain.

Table 3.8 Substrate preferences of CurM ST and OLS ST

	Percent sulfonated[a]	
	CurM ST	OLS ST
(*R*)-3-hydroxydodecanoyl-ACP	73 ± 3	20.4 ± 0.3
(*R*)-3-hydroxytetradecanoyl-ACP	72 ± 2	33 ± 1
(3*R*)-3-hydroxy-5-methoxytetradecanoyl-ACP	74 ± 5	0
(3*S*)-3-hydroxy-5-methoxytetradecanoyl-ACP	28 ± 2	0

[a]Raw HPLC chromatogram peak areas for the ACP-loaded substrate and sulfonated product were used to calculate the fraction of substrate sulfonated. CurM ST reactions were quenched and analyzed after 5 min and OLS ST after 3 h. Data are mean ± standard deviation from triplicate experiments.

Structures of activating STs.

In order to more fully understand the activating STs, crystal structures were solved for CurM ST and OLS ST (Figure 3.8). Both enzymes were crystallized with the product PAP, which was essential for crystal growth. Crystallization of CurM ST required a double-alanine substitution at Gln259-Lys260. The structures define the limits of the ST domains as amino acids 1623–1905 of CurM and 2130–2419 of OLS. Here we use a common amino acid numbering based on the recombinant CurM ST polypeptide, as there are no gaps in the alignment of the CurM and OLS STs.

CurM and OLS STs resemble other STs in their overall architecture, as expected. However, the activating STs are much more similar to each other (RMSD = 0.79 Å for 228 Cα) than either is to the ST of next greatest structural similarity, human heparan sulfate 3-*O*-sulfotransfease (3-OST, RMSD = 1.5 Å for only 114 Cα atoms) (Figure 3.9A,B). In the activating STs, the generic ST core is embellished with 3 additional α-helices (α2–α4) (Figure 3.8, Figure 3.9B).

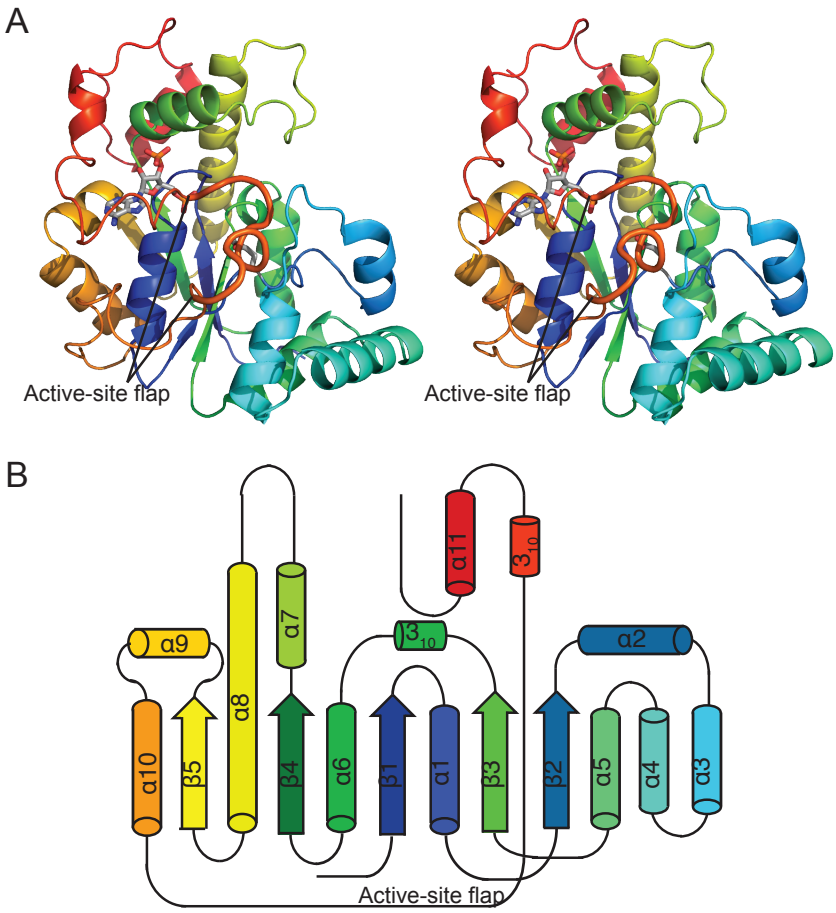


Figure 3.8 Structure of activating STs
A. CurM ST polypeptide. The stereo ribbon diagram is colored as a rainbow from blue at the N-terminus to red at the C-terminus with PAP and Glu60, the proposed catalytic base, in stick (gray C). The active site flap (thick) is labeled. **B.** Topology diagram. CurM ST and OLS ST have an α/β core fold, like other STs, with inserted α-helices (α2-α4) and an extended loop wrapping around the core from α10-α11.

Serendipitous insight into PAPS binding.

PAP binds similarly in CurM ST and OLS ST to a cleft near the C-terminal edge of a central parallel β-sheet (Figure 3.8A, Figure 3.9). As in most other ST structures (except for the 3-OST family) (97) the adenine base is bound in the *anti* conformation. PAP forms an intricate network of hydrogen bonds with amino acids conserved in the activating ST sequences (Figure 3.10A,B Figure 3.2). Some of these interactions occur only in the activating ST family (Glu168, Tyr248, Asp266, Asn268, His272, and backbone interactions with Thr274 and Asp276) while others are common in the STs superfamily (Arg39, Gly41, Ser42, Thr43, Arg161, Ser169, Arg172, Tyr218).

A Zn^{2+} ion from the CurM ST crystallization solution coordinates PAP at a position corresponding to the PAPS sulfate, and provides a serendipitous insight into PAPS binding. The Zn^{2+} coordinates a PAP 5' phosphate oxygen, and three Cl^- ions complete the tetrahedral coordination sphere, resulting in a species with the same charge and shape as sulfate. The Zn^{2+} was identified by anomalous scattering, and the Cl^- ions were placed at the positions of weak anomalous scattering peaks (Figure 3.6). Although $ZnCl_3^-$ is bulkier than SO_3^-, the active site accommodates $ZnCl_3^-$ without perturbing the structure as the surrounding side chains are not significantly shifted relative to their positions in OLS ST (Figure 3.6B). PAPS was modeled into the structure based on the position of $ZnCl_3^-$ (Figure 3.10C, Figure 3.6C). The PAPS sulfate location matches well with the PAPS complex of a distantly related mouse ST (113). Two amino acids, which are conserved in activating STs, are positioned to interact with PAPS sulfate oxygen atoms. Arg39 simultaneously recognizes the PAPS 3'-phosphate and 5'-sulfate. Lys133 forms an ionic interaction with the PAPS 5'-sulfate and hydrogen bonds with amino acids in the active site (Figure 3.10C).

The presence of $ZnCl_3^-$ in the active site implies that Zn^{2+} may inhibit the ST. We tested the activity of CurM ST as a function of Zn^{2+} concentration, and observed a less-than twofold decrease in sulfonation over a range of 1–100 μM (Table 3.1), concentrations far above estimates of free Zn^{2+} in the bacterial cytoplasm (114, 115). Thus, based on these data, we conclude that Zn^{2+} does not interfere with CurM ST activity *in vivo*.

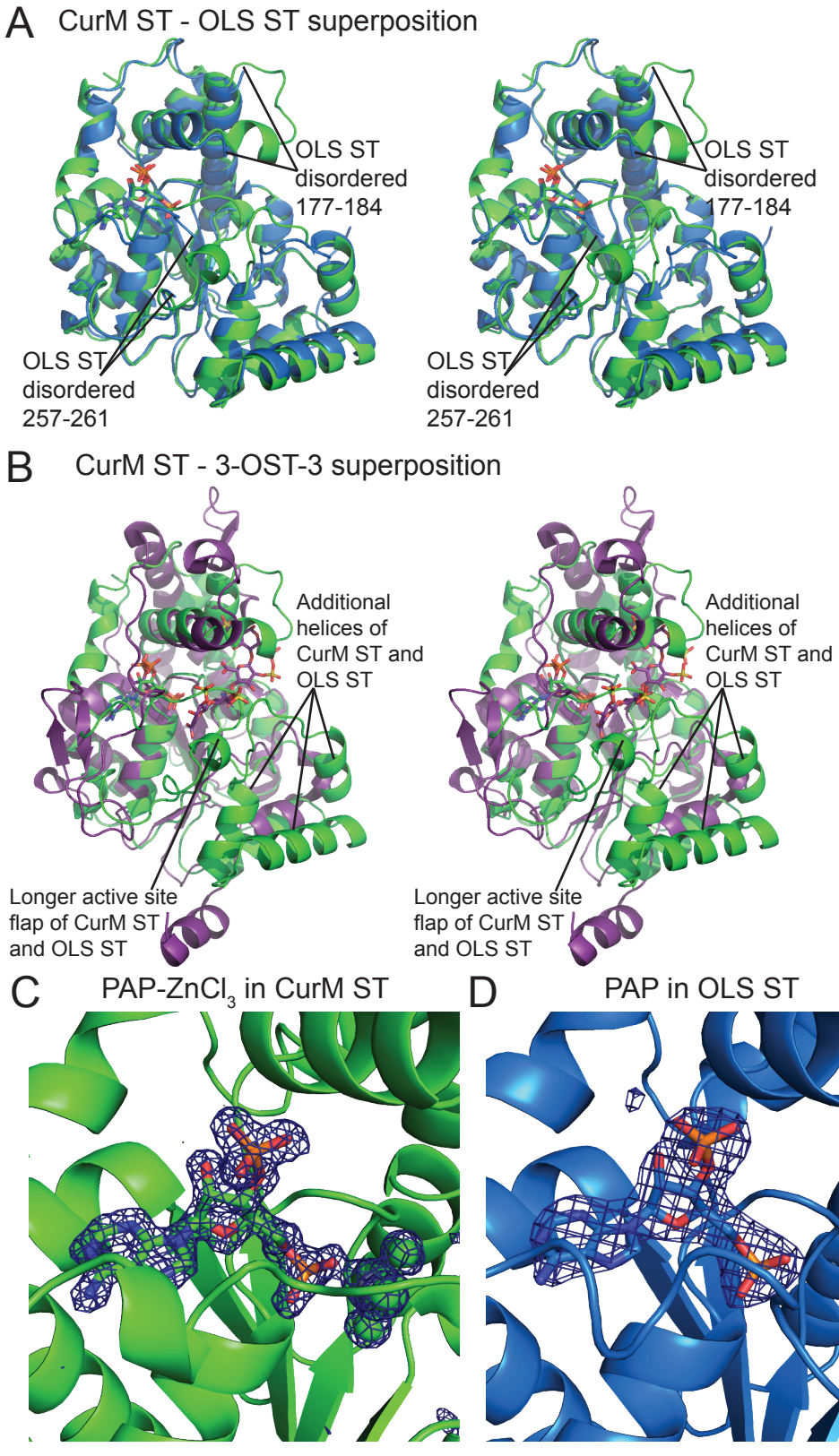

A CurM ST - OLS ST superposition
OLS ST disordered 177-184
OLS ST disordered 257-261
OLS ST disordered 177-184
OLS ST disordered 257-261
B CurM ST - 3-OST-3 superposition
Additional helices of CurM ST and OLS ST
Longer active site flap of CurM ST and OLS ST
Additional helices of CurM ST and OLS ST
Longer active site flap of CurM ST and OLS ST
C PAP-ZnCl3 in CurM ST
D PAP in OLS ST

Figure 3.9 Comparison of ST structures
A. Structure alignment of CurM ST (green) to OLS ST (blue) (RMSD = 0.785 Å for 228 Cα atoms). PAP is shown in stick. **B.** Structure alignment of CurM ST (green) to 3-OST (1T8U (20)) (purple) (RMSD = 1.47 Å for 114 Cα atoms). PAP is shown in stick for CurM ST (green C). PAP and the tetrasaccharide substrate of 3-OST are also shown in stick (purple C). Both parts A and B are shown in stereo. **C.** Electron density of PAP and $ZnCl_3^-$ bound to CurM ST. **D.** Electron density of PAP bound to OLS ST. In both parts C and D, electron density is from an omit Fo-Fc map contoured to 4.0 sigma after removing ligands from the atomic model and running one round of refinement.

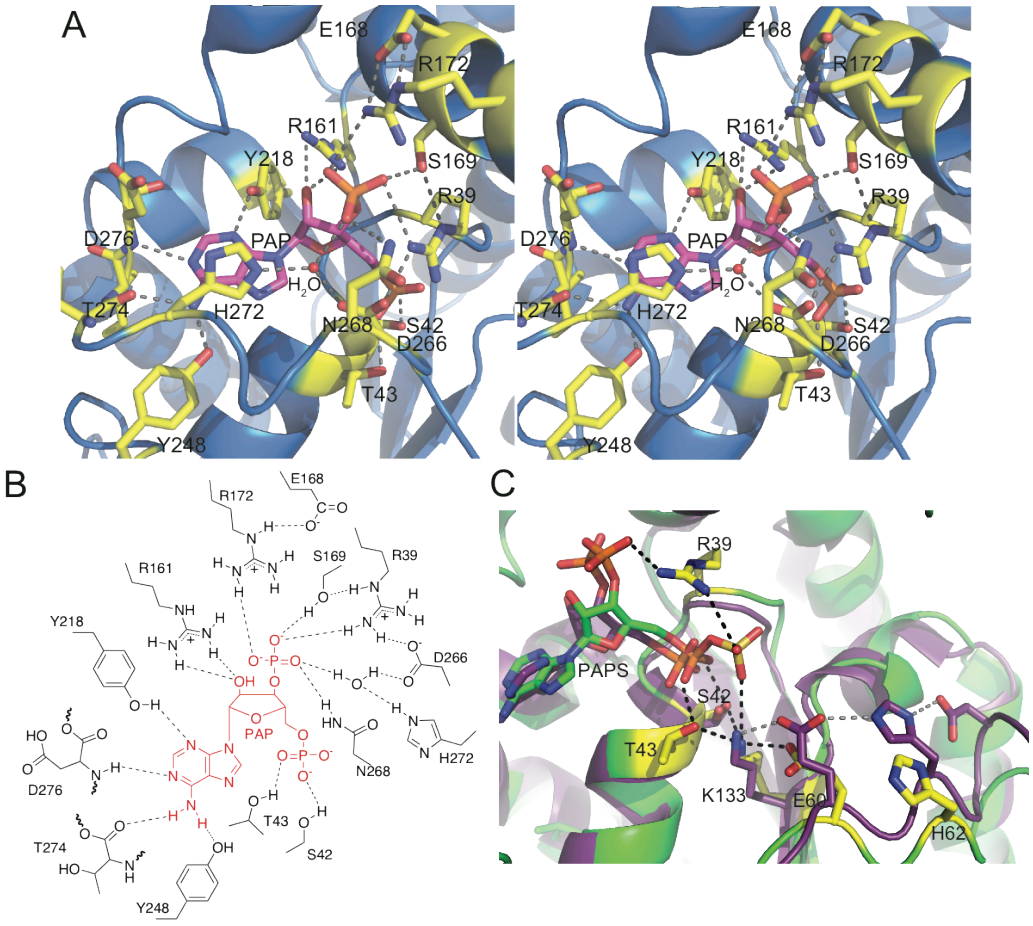


Figure 3.10 Active site of activating STs
A. OLS ST PAP binding site. PAP (magenta C) and surrounding amino acids (yellow C) in an intricate hydrogen-bonding network are shown in stick form in the stereo diagram. **B.** Schematic of PAP binding site. **C.** Comparison of the active sites of CurM ST and 3-OST. Key amino acids are shown in stick form in 3-OST (1T8U) (25) (purple C) and CurM ST (yellow C). Hydrogen bonds in CurM ST and the ionic interaction from Lys133 to PAP are shown in black. Hydrogen bonds in 3-OST are shown in gray. PAPS is modeled into CurM ST from the positions of PAP and $ZnCl_3^-$.

Active site.

Several amino acids near the PAPS sulfate are positioned to affect catalysis (Figure 3.10C). The key residues Glu60 and Lys133 are invariant among the activating STs. Glu60 is well situated to serve as the catalytic base; the PAPS sulfate is in an optimal position for nucleophilic attack by the substrate β-hydroxy group following deprotonation by Glu60 (Figure 3.11). In other STs, the analog of Glu60 is part of a Glu-His-Asp triad linked by hydrogen bonds (110, 116). However, the activating STs lack such a triad. His62 is too distant from Glu60 for a hydrogen bond, and no Asp is present. Lys133 anchors the active site by direct interaction with both Glu60 and the PAPS sulfate. Lys133 is further stabilized by hydrogen bonds with the Ser42 and Thr43 hydroxy groups, which in turn hydrogen bond to PAP 5' phosphate, further linking PAPS to the catalytic residues (Figure 3.10C). Most STs in the structure database have an "anchor" lysine like Lys133.

The importance of these and other amino acids in the active site (Arg39, Thr43, Glu60, His62, Lys133, Ser134) were probed by mutagenesis of CurM ST (Table 3.9). Ala and Gln substitutions at Glu60 reduced activity more than 100-fold, consistent with its proposed role as a catalytic base. A key role was confirmed for the anchor Lys133 (Figure 3.10), for which an alanine substitution reduced activity below detectable levels. Other perturbations of the hydrogen-bonding network through alanine substitutions at Thr43 (linking Lys133 and the PAPS 5' phosphate) and Arg39 (linking the PAPS sulfate to the 3' phosphate) led to a 100-fold reduction in activity. The mutagenesis results are consistent with a model in which Lys133 anchors key players in the active site: the PAPS sulfate, the catalytic base Glu60, and the PAPS 5' phosphate indirectly through Thr43. An Ala substitution at His62 in CurM ST reduced activity only 10-fold, in contrast to an ST with a Glu-His-Asp triad, where the analogous His is essential (110). The activity of this variant suggests His62 may interact with Glu60 during the catalytic cycle but is not necessary for catalysis.

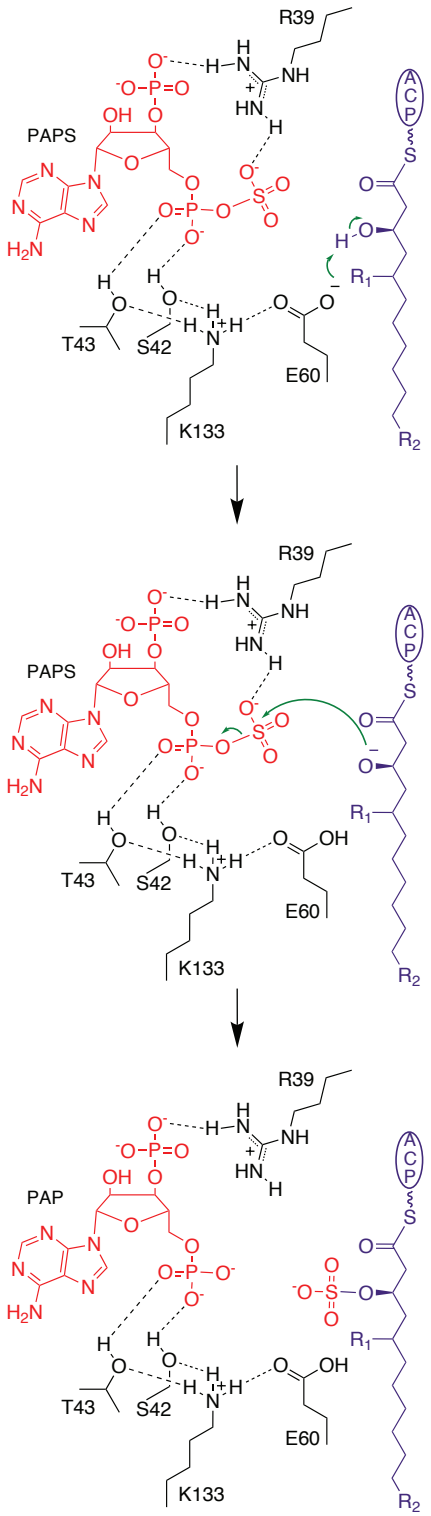


Figure 3.11 Proposed mechanism of catalysis of activating STs

Table 3.9 Activity of CurM ST variants

CurM ST variants	% Activity[a]	Location
Wild type	100	
R39A	1.4 ± 1.1	PAP binding
T43A	1.1 ± 1.5	Active site
E60A	0.3 ± 1.0	Active site
E60Q	0.7 ± 0.6	Active site
H62A	8 ± 1	Active site
K133A	< 0.1	Active site
S134A	52 ± 11	Structural
S261A	131 ± 19	Flap
D266A	1.9 ± 1.1	Flap, PAP binding
D266N	3.5 ± 0.9	Flap, PAP binding
P267A	58 ± 5	Flap
No ST	0	
No PAPS	0	

[a]Raw HPLC chromatogram peak areas for the substrate ((*R*)-3-hydroxymyristoyl-ACP) and product ((*R*)-3-sulfomyristoyl-ACP) were used to calculate the fraction of substrate sulfonated. The activity of each variant was normalized to the wild type within each replicate. Mean ± standard deviation from triplicate experiments is shown.

Dynamic active-site flap.

CurM ST and OLS ST possess a dynamic active-site flap (Figure 3.12A). A 13-residue loop between helices α10 and α11 (residues 255–267) covers the active site in CurM ST and is differently positioned (255, 264–267) and partially disordered (256–263) in OLS ST (Figure 3.12A, Figure 3.13A). The ST superfamily exhibits substantial divergence in length and structure for this region at the entrance to the active site, in accord with the great variety of substrates that are sulfonated (97, 102). However, among the activating STs, the flap sequence is conserved, suggesting that the substrates are similar or that the flaps have a common function. Certainly the dynamic aspect of the CurM and OLS ST flaps hints that they close over the β-hydroxy substrate during catalysis. An adjacent flexible loop (residues 77–82 between α2 and α3) has somewhat different positions in CurM and OLS ST (Figure 3.12A), and may move in conjunction with the active-site flap. Similar order/disorder transitions have been observed in the

mammalian cytosolic STs, where loops and helices surrounding the active site are progressively ordered upon the binding of PAP(S) and substrate (102).

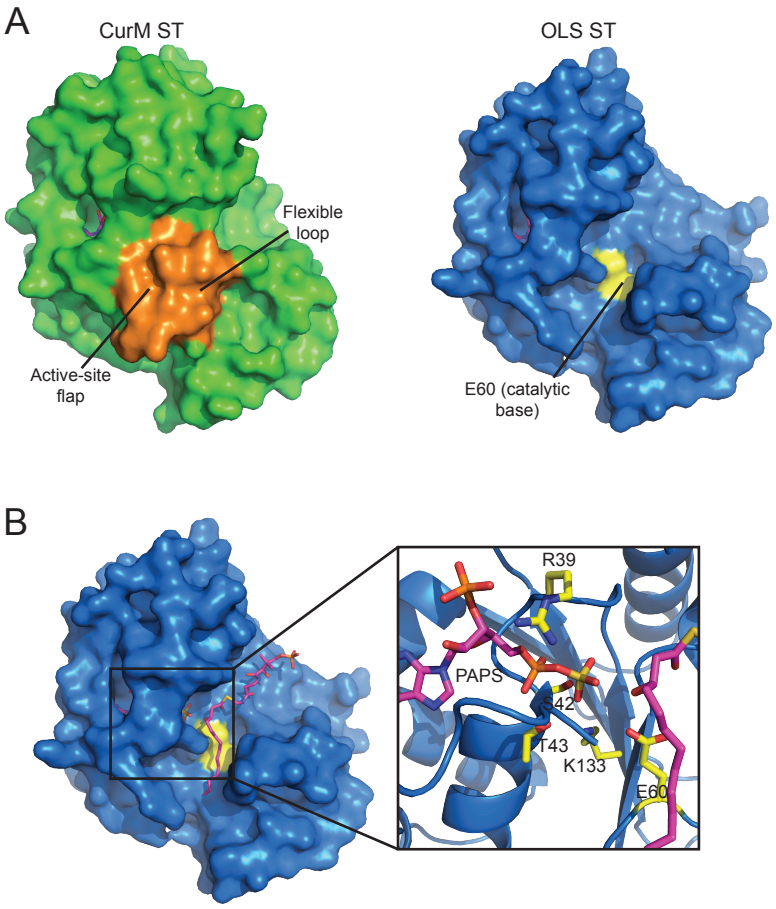


Figure 3.12 Active-site cleft
A. Surface representation of CurM ST (green) and OLS ST (blue). For both proteins PAP is shown in magenta stick and catalytic Glu60 in yellow (not visible in CurM ST). The active-site cleft in OLS ST is exposed due to disorder of the active-site flap. In CurM ST, the active-site flap and flexible loop (orange surfaces) cover the active site but their conformations are influenced by crystallization. **B.** Substrate modeled into OLS ST. (*R*)-3-hydroxymyristoylphosphopantetheine (magenta C, red O, blue N, yellow S, orange P) was modeled into the open substrate-binding cleft of OLS ST.

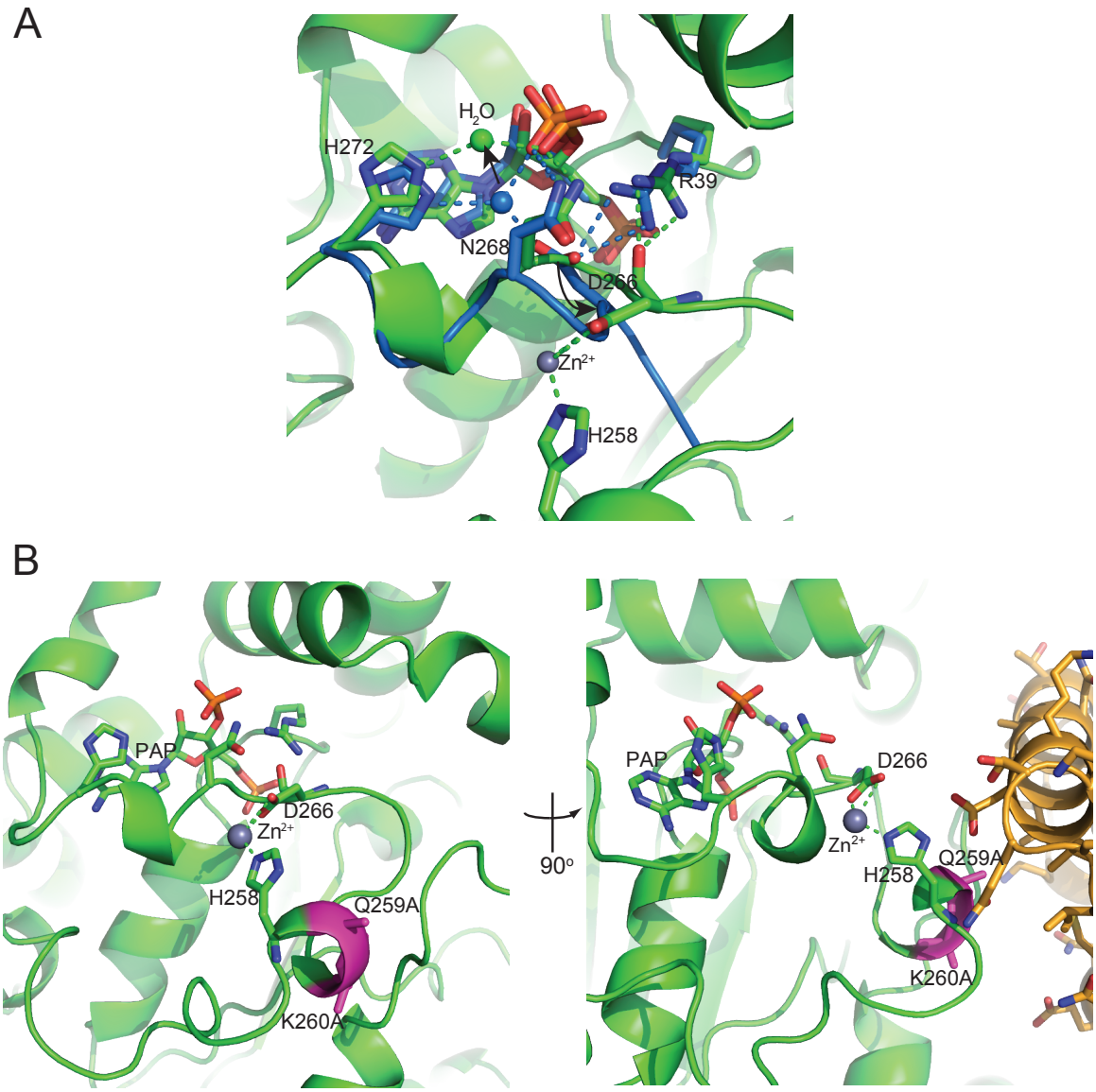


Figure 3.13 Ordered conformation of the active-site flap
A. Differences in CurM ST and OLS ST PAP-binding residues. Residues in CurM ST (green C) and OLS ST (blue C) are shown in stick. Asp266 and His258 coordinate a bound Zn^{2+} ion in CurM ST. In OLS ST, Asp266 points into the active site and contacts Arg39 and an ordered water. In CurM ST, Arg39 instead interacts with a backbone carbonyl. **B.** Position of the surface entropy reduction (SER) mutations that enabled crystallization of CurM ST. SER Ala substitutions are shown as magenta sticks. The crystal contact mediated by the SER mutations is shown (gold C, red O, blue N).

Two features of the CurM ST crystallization contribute to the well-ordered conformation of the active-site flap and flexible loop structure in the absence of substrate. A Zn^{2+} bound on the outside of the protein (not in the PAPS sulfate position) coordinates flap residues His258 and Asp266 (Figure 3.6A, Figure 3.13) and serves to anchor these points of the flap. Asp266 is invariant among activating STs and participates in the PAPS hydrogen bonding network in the OLS ST structure through a water-bridged hydrogen bond with the PAP 3'-phosphate and a direct hydrogen bond with Arg39 (Figure 3.13A). The alanine substitutions to facilitate crystallization of CurM ST (Q259A/K260A) were at non-conserved sites in the active-site flap, and also contribute to trapping the flap in the observed conformation. Q259A/K260A mediates a crystal contact that also includes Tyr80 and Leu81 from the adjacent flexible loop (Figure 3.13B). Together the crystal contact, the alanine substitutions, and the surface Zn^{2+} ion trapped the CurM ST flap in an ordered/closed conformation, however, this particular ordered conformation may not be relevant to ST function.

Substrate binding and specificity.

We modeled substrates into the active sites in positions to interact with both the PAPS sulfate and Glu60, the proposed catalytic base. The active-site flap was ignored in the modeling because it is unknown how the flap closes over the substrate during catalysis. The structure of a ternary complex of 3-OST with PAP and a tetrasaccharide substrate [1T8U (110)] was used as a guide to place the β-hydroxy groups of the OLS ST substrate, (*R*)-3-hydroxymyristoyl-phosphopantethiene, and the curacin PKS ST substrate mimic, (3*R*)-3-hydroxy-5-methoxymyristoyl-phosphopantethiene (Figure 3.12B). The phosphopantetheine arm and alkyl chain are well matched to the length of the substrate channel when the substrate β-hydroxy group is adjacent to Glu60 and to the PAPS sulfate.

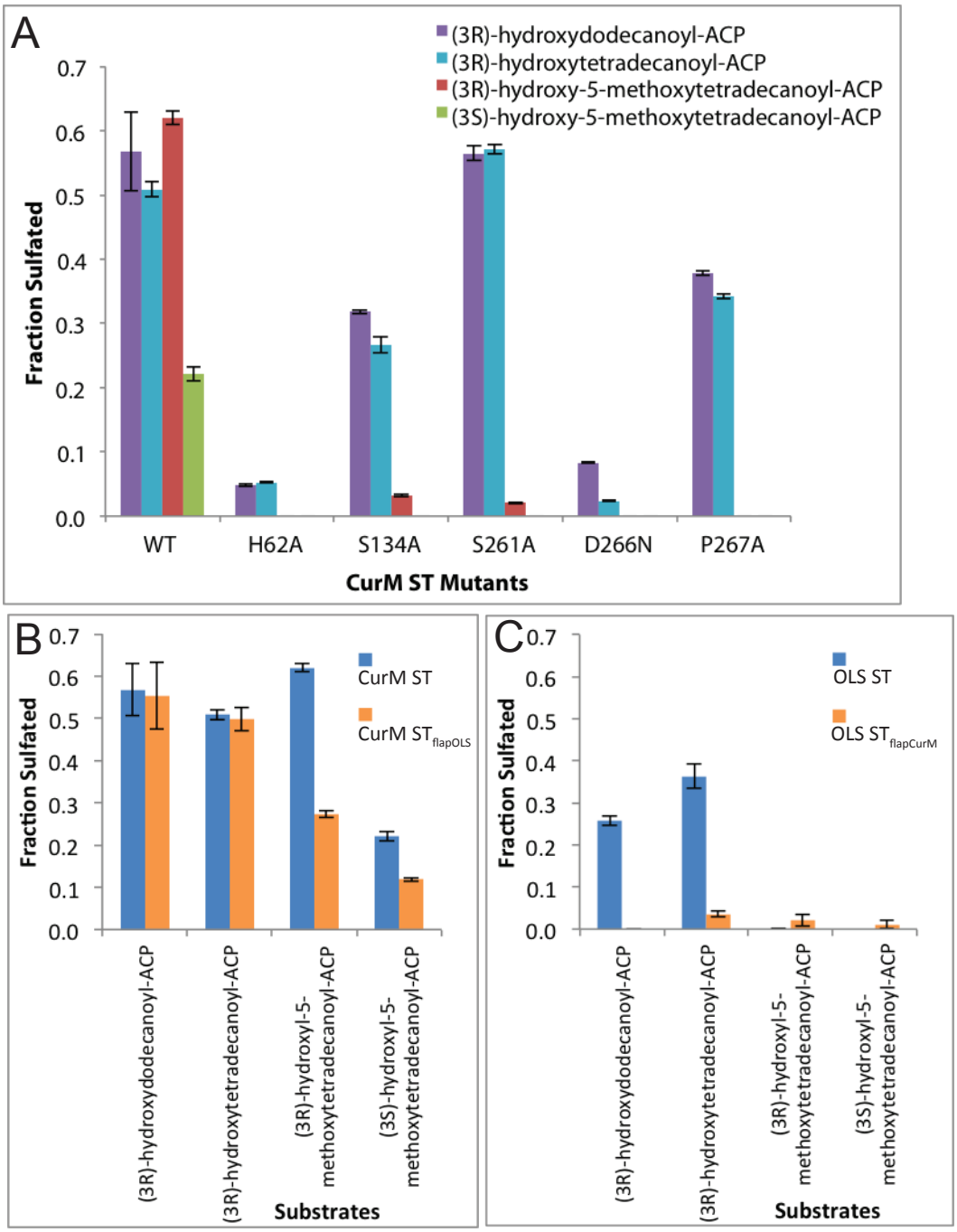

A
(3R)-hydroxydodecanoyl-ACP
(3R)-hydroxytetradecanoyl-ACP
(3R)-hydroxy-5-methoxytetradecanoyl-ACP
(3S)-hydroxy-5-methoxytetradecanoyl-ACP
Fraction Sulfated
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WT
H62A
S134A
S261A
D266N
P267A
CurM ST Mutants
B
CurM ST
CurM ST flapOLS
Fraction Sulfated
(3R)-hydroxydodecanoyl-ACP
(3R)-hydroxytetradecanoyl-ACP
(3R)-hydroxyl-5-methoxytetradecanoyl-ACP
(3S)-hydroxyl-5-methoxytetradecanoyl-ACP
Substrates
C
OLS ST
OLS ST flapCurM
Fraction Sulfated
(3R)-hydroxydodecanoyl-ACP
(3R)-hydroxytetradecanoyl-ACP
(3R)-hydroxyl-5-methoxytetradecanoyl-ACP
(3S)-hydroxyl-5-methoxytetradecanoyl-ACP
Substrates

Figure 3.14 Substrate preference for ST variants
A. CurM ST single amino acid substitutions. Raw HPLC chromatogram peak areas for the substrate and sulfonated product after 5-min reactions were used to calculate the fraction of substrate sulfonated. Mean ± standard deviation from duplicate experiments is shown. **B.** Substrate preference for CurM ST chimera with OLS ST flap. Experiment was performed as in A. Mean ± standard deviation from triplicate experiments is shown. CurM $ST_{flapOLS}$ shows similar activity to CurM ST with simple alkyl substrates but decreased activity with C5-methoxy substrates. **C.** Substrate preference for OLS ST chimera with CurM ST flap. Raw HPLC chromatogram peak areas for the substrate and sulfonated product after 3 h reactions were used to calculate the fraction of substrate sulfonated. Mean ± standard deviation from triplicate experiments is shown. OLS $ST_{flapCurM}$ shows very low overall activity.

The modeling did not provide an explanation for the OLS ST substrate selectivity, as substrates both with and without the C5-methoxy were easily modeled into both enzymes. Thus we probed the function of the active-site flap with two types of site-directed mutagenesis experiments. Single-site substitutions were made to three conserved amino acids in the active-site flap of CurM ST (Ser261, Asp266, Pro257) (Table 3.9). Activity with a simple alkyl substrate was reduced nearly 100-fold when Asp266 was substituted by Ala (2% of WT) or Asn (4% of WT), consistent with the Asp266 interaction with PAPS through Arg39 and an ordered water molecule (Figure 3.10A,B Figure 3.13A). The P267A and S261A variants had modestly reduced activity relative to the wild type for a simple alkyl substrate. We then tested the variants with the greatest activity (H62A, S134A, S261A, D266N, P267A) with all four substrates (Figure 3.14A). Remarkably, alanine substitutions at conserved flap residues Ser261 and Pro267 resulted in substantial selection toward substrates lacking the C5-methoxy group, similar to OLS ST. This unexpected result suggests that the structure of the closed active-site flap affects substrate selectivity and that small changes to the flap sequence affect its structure.

The other mutagenesis experiments more directly tested the role of the active-site flap in substrate specificity. Chimeric proteins were created in which the CurM ST and OLS ST flaps (residues 254–270) were exchanged. The activity of CurM $ST_{flapOLS}$ with C5-methoxy substrates decreased two-fold relative to wild type while the activity with the simple alkyl substrates was unchanged (Figure 3.14B). This resulted in a CurM ST with activity more like that of OLS ST and demonstrates the importance of the flap in

substrate selectivity. However, the flap is not the sole determinant of substrate selectivity, as OLS $ST_{flapCurM}$ had negligible activity for all substrates (Figure 3.14C).

Discussion

Activating sulfotransferase family.

CurM ST and OLS ST are prototypes of a family of functional-group-activating sulfotransferases in complex metabolic pathways. Both CurM ST and OLS ST catalyze β-hydroxy group sulfonation, the first step in this new mode of metabolite offloading and termination. Sulfonation activates the substrate for the unprecedented TE-catalyzed decarboxylation and sulfate elimination (27), and hence we have named the family "activating" sulfotransferases. The activating ST structure is most similar to the mammalian heparan sulfate D-glucosaminyl 3-*O*-sulfotransferase (3-OST) family of STs. Although the sequence identity is low (<15%), activating STs and 3-OST family members assist catalysis using a common set of amino acids. Results of site-directed mutagenesis of CurM ST are consistent with a catalytic mechanism in which Glu60 is the catalytic base and Lys133 anchors the PAPS sulfate and Glu60.

The sequence database contains several other uncharacterized gene products of high similarity to CurM ST (31%-43% identity) that occur in microbial natural product biosynthetic pathways without ACP-ST-TE tridomains (Figure 3.3). These STs are not expected to activate hydroxy groups, but instead to produce sulfated natural products. The catalytic residues and active-site flap of the activating STs are conserved in these similar but "non-activating" ST sequences, indicative of recent common ancestry (Figure 3.3). Natural products pathways do not include exclusively STs of the activating ST family. For example the glycopeptide ST Teg12 (43) (16% identity to CurM ST) clearly belongs to a different ST family.

OLS and PKS STs display different substrate selectivity.

The assay results for OLS ST are the first biochemical studies of a domain from an olefin synthase. Consistent with genetic deletion and feeding studies (52), OLS ST acted on substrates with long alkyl chains. However, it had no detectable activity with C5-methoxy substrates in contrast to the CurM ST, which was active with all substrates tested. The exclusion of C5-methoxy substrates by OLS ST is presumably due to a feature of the closed active-site flap because the OLS ST and CurM ST surfaces "under"

the active-site flap are identical in charge and structure. The flap-specificity hypothesis is consistent with the activity of the CurM $ST_{flapOLS}$ chimera, which had decreased activity with C5-methoxy substrates but similar activity to CurM ST with the simple alkyl substrates. The active-site flap also contributes to substrate selectivity in other ST sub-families. For example, the human ST SULT2A1 active-site flap controls substrate selectivity by limiting the size of substrates that can access the active site after PAPS binding (117). Although the activating ST flap is highly diverged in sequence from the flaps of other STs such as SULT2A1, it appears to be used broadly within the ST superfamily to control substrate specificity.

CurM ST had only a weak preference for the substrate mimic with the natural (*R*)-β-hydroxy stereochemistry over the non-natural (*S*) (Table 3.8, (27)), consistent with the idea that the ST is likely to encounter only substrates with the natural stereochemistry. The stereochemistry is predicted by sequence analysis of the upstream KR domain, which specifies an (*R*)-β-hydroxy product for both STs (118). Under our assay conditions, CurM ST is a much faster enzyme than OLS ST (Figure 3.14B,C), even for substrates linked to the cognate OLS ACP (43% identity to CurM ACP). OLS ST may be more reactive with the natural substrate (6–8 carbons longer than the test substrates) or within the context of the full-length OLS polypeptide where substrates are linked to a fused ACP.

Distinctive structural features of activating STs.

We visualized the PAPS sulfate in CurM ST from the serendipitous binding of $ZnCl_3^-$ in the sulfate position (119, 113). The PAPS sulfate is thus deduced to be less than 5 Å from the putative catalytic base, Glu60. The β-hydroxy group of the modeled substrate is in a position to be activated by Glu60 and subsequently to attack the sulfate of PAPS to complete sulfonate transfer (Figure 3.12B). Arg39 appears key in distinguishing PAPS from other nucleotides, as it simultaneously recognizes the 3'-phosphate and the 5'-sulfate. Additionally, four hydrogen bonds provide specificity for the adenine base (Figure 3.10A).

The structures of CurM ST and OLS ST together provide evidence of a dynamic active-site flap distinct to activating STs (Figure 3.12A). In the OLS ST, the flap is disordered, exposing an obvious substrate-binding cleft adjacent to the PAPS sulfate (Figure 3.12B). Crystallization of CurM ST captured a closed/ordered flap, as expected

to follow substrate binding. However, no substrate was present, and a bound Zn^{2+}, the two alanine substitutions, and a crystal contact contributed to the observed flap structure. The conservation of the length and amino acid sequence of the flap highlights its importance, as does the reduced activity of substitutions at invariant Asp266 in the flap. Asp266 is proposed to guide the flap into the proper closed position by interacting with PAPS, thereby ensuring that PAPS is bound before the flap closes. Other residues conserved in the flap of activating STs (Pro267, Ser261) are involved directly or indirectly in substrate recognition, as Ala variants altered the substrate specificity, an unexpected finding for amino acids conserved in the Cur and OLS enzymes. Flap chimeras resulted in a CurM ST with substrate specificity more like OLS ST but an OLS ST with negligible activity. Thus the flap is not the sole determinant of substrate selectivity, and we infer that flap-core interactions also modulate and control the substrate selectivity of the activating STs.

Functional-group activation.

CurM ST and OLS ST carry out an activating sulfonation that creates a favorable leaving group to assist subsequent decarboxylation; the transient sulfate is present on an intermediate but not in the final product. The curacin A pathway also contains an "activating" halogenase, which functionalizes a carbon by chlorination for subsequent cyclopropane formation (22, 23). A similar chlorination strategy is employed to form a cyclopropane in the biosynthesis of coronamic acid (120). Activation of a hydroxy group by phosphorylation with phosphate, pyrophosphate or AMP is common in primary and secondary metabolism. The similarity of activating sulfonation and phosphorylation is striking in mevalonate-5-diphosphate decarboxylase (MDD) in the cholesterol biosynthetic pathway. MDD activates a β-hydroxy group by phosphorylation to facilitate decarboxylation and generate the terminal double bond of isopentenyl diphosphate (60, 27), chemically analogous to ST-TE catalysis. Unlike MDD, the ACP-ST-TE tridomain has the additional task of thioester hydrolysis. For the thioesterase-decarboxylase function, nature adapted the common thioesterase of PKS and NRPS systems and co-opted a sulfotransferase to facilitate the decarboxylation step. Activation by sulfonation is a rare variation on the theme of leaving-group activation by phosphorylation common

to other metabolic pathways, and clearly is an effective strategy to activate substrates for combined thioester hydrolysis and decarboxylation.

Our structure-function studies define a distinct family of STs where most members have a surprising role in hydroxy group activation. This family includes ST domains from biosynthetic systems that produce natural products (polyketides, such as curacin A, and nonribosomal peptides) and hydrocarbons (OLS). The CurM ST and OLS ST structures reveal active site residues, details of PAPS binding, and a dynamic active-site flap, which was shown to influence substrate specificity. The CurM and OLS STs expand our view of nature's remarkable synthetic toolbox and highlight promising engineering applications. The OLS ST is applicable to liquid biofuel production as it leads directly to long-chain hydrocarbons. The CurM ST, which accepts substrates containing a variety of functional groups, also holds value for use in natural product structural diversification.

Author contributions to the manuscript

JGM, JLS, and DHS designed the research. PW contributed new reagents (CoA substrates) synthesized by SK. LG performed the phylogenetic analysis. JGM performed all other experiments. JGM, EBE, and JLS analyzed the data. JGM and JLS wrote the paper. JGM, EBE, LG, PW, WHG, DHS and JLS revised and edited the paper.