BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

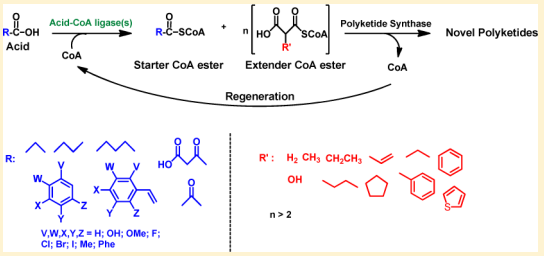
Establishing a Toolkit for Precursor-Directed Polyketide Biosynthesis: Exploring Substrate Promiscuities of Acid-CoA Ligases

Maybelle Kho Go, Jeng Yeong Chow, Vivian Wing Ngar Cheung, Yan Ping Lim, and Wen Shan Yew*

Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597

Supporting Information

ABSTRACT: Polyketides are chemically diverse and medicinally important biochemicals that are biosynthesized from acyl-CoA precursors by polyketide synthases. One of the limitations to combinatorial biosynthesis of polyketides has been the lack of a toolkit that describes the means of delivering novel acyl-CoA precursors necessary for polyketide biosynthesis. Using five acid-CoA ligases obtained from various plants and microorganisms, we biosynthesized an initial library of 79 acyl-CoA thioesters by screening each of the acid-CoA ligases against a library of 123 carboxylic acids. The library of acyl-CoA thioesters includes derivatives of cinnamyl-CoA, 3-phenylpropanoyl-CoA, benzoyl-CoA, phenylacetyl-CoA, malonyl-CoA, saturated and unsaturated aliphatic CoA thioesters, and bicyclic aromatic CoA thioesters. In our search for the biosynthetic routes of novel acyl-CoA precursors, we discovered two previously unreported malonyl-CoA derivatives (3-thiophenemalonyl-CoA and phenylmalonyl-CoA) that cannot be produced by canonical malonyl-CoA synthetases. This report highlights the utility and importance of determining substrate promiscuities beyond conventional substrate pools and describes novel enzymatic routes for the establishment of precursor-directed combinatorial polyketide biosynthesis.



Polyketides are structurally diverse secondary metabolites produced by microorganisms and plants that are used for a variety of therapeutic purposes, including their use as antimicrobial, antifungal, and anticancer agents.[1,2] They are biosynthesized from acyl-CoA precursors by polyketide synthases (PKSs) and, because of their chemical complexity, are not easily synthesized and structurally manipulated by chemical means. There are three types of PKSs.[3] Type I PKSs are modular, multifunctional enzymes that catalyze noniterative polyketide-chain biosynthesis. Type II PKSs are nonmodular, multifunctional enzymes that iteratively catalyze the biosynthesis of the polyketide chain. Type III PKSs are acyl-carrier protein-independent enzymes that catalyze the iterative biosynthesis of the polyketide chain within a single active site. Irrespective of the mechanism of polyketide-chain extension, PKSs rely on acyl-CoA precursors for polyketide biosynthesis.

Considerable effort has been spent in exploring the biosynthesis of these biochemicals,[4,5] and although there has been excellent progress,[6−8] rationally guided combinatorial biosynthesis, either in vitro or in vivo, to generate novel and biomedically relevant polyketides remains unattainable. One of the contributing factors to this unrealized biosynthetic endeavor is the lack of a toolkit that describes the means of delivering novel acyl-CoA precursors necessary for polyketide biosynthesis. In this study, we report the substrate range of five acid-CoA ligases and the corresponding library of 79 acyl- and aryl-CoAs that can be biosynthesized using these enzymes for precursor-directed polyketide biosynthesis, providing enzymatic routes to the biosynthesis of novel polyketides.

Aryl- and acyl-CoA precursors can be biosynthesized by a large family of adenylate-forming enzymes that catalyze a two-step adenylation and thioester-forming reaction.[9−11] These acid-CoA ligases catalyze acid-CoA thioester formation through the adenylation of the carboxylic acid substrate. There are a number of substrate specificity studies on acid-CoA ligases.[12−17] These studies are often limited to structural derivatives of the cognate substrate, focusing on derivatives of malonic, cinnamic, 3-phenylpropanoic, phenylacetic, and aromatic acids for their respective acid-CoA ligases. As such, an exploration of substrate promiscuities across the different carboxylic acid classes has not been systematically attempted or reported.

In this study, we determined the extent of substrate promiscuity of five acid-CoA ligases. These enzymes were chosen on the basis of the availability of genomic DNA and the reported and/or expected substrate range; 123 carboxylic acids, such as malonates, cinnamates, 3-phenylpropanoates, phenylacetates, benzoates, saturated and unsaturated aliphatics, naphthalene-, pyridine-, and quinoline-carboxylates, and their derivatives, were used in the study. Our study of malonyl-CoA synthetase from *Rhizobium trifolii* (RtMCS) corroborated previous findings[14] that the acid-CoA ligase was promiscuous toward a range of malonic acid derivatives ranging from methylmalonic acid to isopropylmalonic acid to allymalonic

Received: April 3, 2012
Revised: May 10, 2012
Published: May 15, 2012

acid; in addition, we report the enzymatic synthesis of hydroxymalonyl-CoA by RtMCS (the biosynthesis of hydroxymalonyl-CoA by MatB from *Streptomyces coelicolor* was first reported in 2011 by Keatinge-Clay and co-workers[18]). A previous report on the promiscuity of cinnamyl-CoA ligase from *S. coelicolor* (CCL) was limited to cinnamic acid derivatives;[12] in this study, we report the biosynthesis of phenylpropanoyl-CoA and its derivatives, as well as short alkyl-CoAs such as butanoyl-, pentanoyl-, and hexanoyl-CoAs by CCL.

4-Coumarate-CoA ligase from *Nicotiana tabacum* (Nt4CL) and benzoate-CoA ligase from *Rhodopseudomonas palustri* (RpBZL) are two previously reported acid-CoA ligases[19] that use cinnamic acid and benzoic acid substrates (and their derivatives), respectively. In addition to cinnamate acid derivatives, we found that Nt4CL could also use 3-phenylpropanoic acid (and its derivatives), medium- to long-chain saturated alkylic acids (hexanoic, heptanoic, octanoic, nonanoic, and decanoic acids), and naphthalenecarboxylic acids and quinolinecarboxylic acids as substrates for the biosynthesis of the corresponding acyl-CoA thioesters. Apart from a wide range of benzoic acid derivatives, we report the biosynthesis of short-chain saturated and unsaturated alkyl-CoAs (such as propanoyl-CoA, pentanoyl-CoA, and *trans*-2-butenoyl-CoA or crotonoyl-CoA, respectively) by RpBZL, providing novel enzymatic routes for polyketide biosynthesis efforts.

A promiscuous phenylacetate-CoA ligase (PCL) was previously reported from the fungus *Penicillium chrysogenum* by Janssen and co-workers.[17] In this study, we report the substrate range of a functionally orthologous PCL from *S. coelicolor* (20% identical amino acid sequence): apart from the previously reported catalysis of phenylacetic acid (and its derivatives) and saturated aliphatic acids, we report the biosynthesis of unsaturated alkyl-CoAs and two novel malonyl-CoA derivatives by PCL (3-thiophenemalonyl-CoA and phenylmalonyl-CoA). It is noteworthy that given the broad substrate range of RtMCS toward malonic acid derivatives, it could not catalyze the production of 3-thiophenemalonyl-CoA and phenylmalonyl-CoA. This study highlights the utility and importance of determining substrate promiscuities beyond conventional substrate pools and provides an additional tool(s) for the establishment of precursor-directed combinatorial polyketide biosynthesis.

MATERIALS AND METHODS

The carboxylic acid substrates for acid-CoA ligases were purchased from Sigma-Aldrich Co. (St. Louis, MO), Tokyo Chemical Industry (TCI) Co., Extrasynthese Co., and Lier Chemical Co. (Sichuan, China). A total of 123 carboxylic acids were tested: cinnamate type, cinnamic acid, 2-fluorocinnamic acid, 3-fluorocinnamic acid, 4-fluorocinnamic acid, α-fluorocinnamic acid, 2-chlorocinnamic acid, 3-chlorocinnamic acid, 4-chlorocinnamic acid, 4-methylcinnamic acid, α-methylcinnamic acid, 2-hydroxycinnamic acid, 4-hydroxycinnamic acid, 2-methoxycinnamic acid, 4-methoxycinnamic acid, 2,4-dimethoxycinnamic acid, 3,4-dimethoxycinnamic acid, 3-chloro-4-methoxycinnamic acid, 3-methoxy-4-hydroxycinnamic acid (ferulic acid), 4-hydroxy-3,5-dimethoxycinnamic acid, and 4-nitrocinnamic acid; phenylpropanoate type, 3-phenylpropanoic acid, 2-hydroxy-3-phenylpropanoic acid, 3-(2-bromophenyl)-propanoic acid, 3-(2-methoxyphenyl)propanoic acid, 3-(3-chloro-4-methoxyphenyl)propanoic acid, 3-(3-chlorophenyl)-propanoic acid, 3-(3-methoxyphenyl)propanoic acid, 3-(3,5-dichlorophenyl)propanoic acid, 3-(3,5-difluorophenyl)-propanoic acid, 3-(4-fluorophenyl)propanoic acid, 3-(4-methoxyphenyl)propanoic acid, 3-(4-hydroxyphenyl)-2-hydroxypropanoic acid, and 3-(3,4-dihydroxyphenyl)propanoic acid; benzoate type, benzoic acid, 2-fluorobenzoic acid, 3-fluorobenzoic acid, 4-fluorobenzoic acid, 2-chlorobenzoic acid, 3-chlorobenzoic acid, 4-chlorobenzoic acid, 2-bromobenzoic acid, 3-bromobenzoic acid, 4-bromobenzoic acid, 2-iodobenzoic acid, 3-iodobenzoic acid, 4-iodobenzoic acid, 2-aminobenzoic acid, 4-aminobenzoic acid, 2-acetoxybenzoic acid (acetylsalicylic acid), 2-hydroxy-3,5-diiodobenzoic acid, 2-hydroxybenzoic acid (salicylic acid), 2-methoxybenzoic acid, 2-methylbenzoic acid (*o*-toluic acid), 2,3-dihydroxybenzoic acid, 2,4-dihydroxybenzoic acid, 2,6-difluorobenzoic acid, 2,6-dimethylbenzoic acid, 2.5-dihydroxybenzoic acid, 3-aminobenzoic acid, 3,4-dimethoxybenzoic acid (veratric acid), 3,4,5-trihydroxybenzoic acid (gallic acid), 3,5-dihydroxy-4-methoxybenzoic acid, 3,5-dihydroxybenzoic acid, 3,5-dimethylbenzoic acid, and 4-amino-2-hydroxybenzoic acid (*p*-aminosalicylic acid); phenylacetate type, phenylacetic acid, phenoxyacetic acid, 2,4-dichlorophenoxyacetate, 4-fluorophenoxyacetic acid, 4-hydroxyphenylacetic acid, phenylpyruvic acid, 4-methoxyphenylacetate, and 2-hydroxyphenylacetic acid; naphthalene, pyridine, and quinoline type, 1-naphthalenecarboxylic acid, 2-naphthalenecarboxylic acid, pyridine-2-carboxylic acid (picolinic acid), pyridine-3-carboxylic acid (nicotinic acid), 2-chloropyridine-3-carboxylic acid, pyrazine carboxylic acid, pyridine-4-carboxylic acid (isonicotinic acid), 2-quinolinecarboxylic acid (quinaldic acid), 2-quinoxalinecarboxylic acid, 3-quinolinecarboxylic acid, 4-quinolinecarboxylic acid, and isoquinoline-1-carboxylic acid; saturated aliphatic type, acetic acid, propanoic acid, butanoic acid, valeric acid (pentanoic acid), hexanoic acid, heptanoic acid, octanoic acid, nonanoic acid, decanoic acid, acetoacetic acid, pimelic acid, pyruvic acid, and succinic acid; unsaturated aliphatic type, 2-hexenoic acid, 2-methyl-2-butenoic acid (tiglic acid), 2-methyl-2-propenoic acid (methacrylic acid), 2-pentenoic acid, 2-propenoic acid (acrylic acid), 3-butenoic acid, 3-hexenoic acid, 3-methyl-2-butenoic acid (3,3-dimethylacrylic acid), 3-methyl-4-pentenoic acid, 3-pentenoic acid, 4-pentenoic acid, 5-hexenoic acid, and *trans*-2-butenoic acid (crotonic acid); malonic type, malonic acid, methylmalonic acid, phenylmalonic acid, 3-thienylmalonic acid, allylmalonic acid, benzylmalonic acid, butylmalonic acid, cyclopentylmalonic acid, ethylmalonic acid, hydroxymalonic acid, isopropylmalonic acid, and shikimic acid.

The chemical structures of all substrates tested are shown in Figure S1 of the Supporting Information. The coupling enzymes adenylate kinase (AK), pyruvate kinase (PK), and lactic dehydrogenase (LDH) and 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) were purchased from Sigma-Aldrich Co. All reagents were the highest quality grade commercially available.

Cloning, Expression, and Protein Purification of Phenylacetate-CoA Ligase (PCL) from *S. coelicolor* A3(2). The gene encoding PCL (GI|1099823) was amplified via polymerase chain reaction (PCR) from genomic DNA isolated from *S. coelicolor* A3(2) (ATCC) using Platinum *Pfx* DNA polymerase (Invitrogen). The PCR mixture (100 μL) contained 1 ng of plasmid DNA, 10 μL of 10× *Pfx* amplification buffer, 1 mM $MgSO_4$, dNTPs (0.4 mM each), 40 pmol of each primer (forward primer 5′-CGAAGGAGTCCATATGAGCAGCGAGCCGACGACCGGGACGGCC-3′ and reverse primer 5′-GTCCGTGCTCTCGAGTCACGCGCCCCGCTGGTCCCACACCCG-3′), and 5 units

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568−4579

of Platinum *Pfx* DNA polymerase. The gene was amplified using a PTC-0200G Thermal Cycler (Bio-Rad Laboratories), with the following parameters: 94 °C for 2 min followed by 30 cycles of 94 °C for 1 min, 60 °C for 1 min and 15 s, and 68 °C for 3 min, and a final extension of 68 °C for 10 min. The amplified gene was cloned into a modified pET-15b vector (Novagen) in which the N-terminus contained 10 His residues (kindly provided by J. Gerlt, University of Illinois, Urbana, IL).[20]

The protein was expressed in *Escherichia coli* strain BL21(DE3). Transformed cells were grown at 37 °C in LB broth (supplemented with 100 μg/mL ampicillin) to an OD_{600} of 0.6, and isopropyl D-thiogalactopyranoside (IPTG) (0.1 mM) was added to induce protein expression for 16 h. The cells were harvested by centrifugation and resuspended in binding buffer [5 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl (pH 7.9)] and lysed by sonication. The lysate was cleared by centrifugation, and the His-tagged protein was purified using a column of chelating Sepharose Fast Flow (GE Healthcare Bio-Sciences Corp.) charged with Ni^{2+} ion. The cell lysate was applied to the column in binding buffer, washed with buffer containing 154 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl (pH 7.9), and eluted with 100 mM L-histidine, 0.5 M NaCl, and 20 mM Tris-HCl (pH 7.9). The N-terminal His tag was removed with thrombin (GE Healthcare Bio-Sciences Corp.) according to the manufacturer's instructions, and the proteins were purified to homogeneity on a Q Sepharose High Performance column (GE Healthcare Bio-Sciences Corp.) equilibrated with binding buffer [25 mM Tris-HCl (pH 7.9)] and eluted with a linear gradient from 0 to 0.5 M elution buffer [1 M NaCl and 25 mM Tris-HCl (pH 7.9)].

Cloning, Expression, and Protein Purification of Malonyl-CoA Synthetase (RtMCS) from *R. trifolii*. The gene encoding RtMCS (GI:3982573) was amplified via PCR from genomic DNA isolated from *R. trifolii* (ATCC) using Platinum *Pfx* DNA polymerase (Invitrogen). The PCR mixture (100 μL) contained 1 ng of plasmid DNA, 10 μL of 10× *Pfx* amplification buffer, 1 mM $MgSO_4$, dNTPs (0.4 mM each), 40 pmol of each primer (forward primer 5′-GAGGAGGGCGCATATGAGCAACCATCTTTTCGACG-3′ and reverse primer 5′-GAGAGCGCGGGATCCTTACGTCCTGGTATAAAGATC-3′), and 5 units of Platinum *Pfx* DNA polymerase. The gene was amplified using a PTC-0200G Thermal Cycler (Bio-Rad Laboratories), with the following parameters: 94 °C for 2 min followed by 40 cycles of 94 °C for 1 min, 55 °C for 1 min and 15 s, and 68 °C for 3 min, and a final extension of 68 °C for 10 min. The amplified gene was cloned into a modified pET-15b vector (Novagen) in which the N-terminus contained 10 His residues, as previously described. The protein was expressed in *E. coli* strain BL21(DE3). Transformed cells were grown at 25 °C in LB broth (supplemented with 100 μg/mL ampicillin) to an OD_{600} of 0.6, and IPTG (0.1 mM) was added to induce protein expression for 24 h. RtMCS was purified from the harvested cells as previously described for PCL.

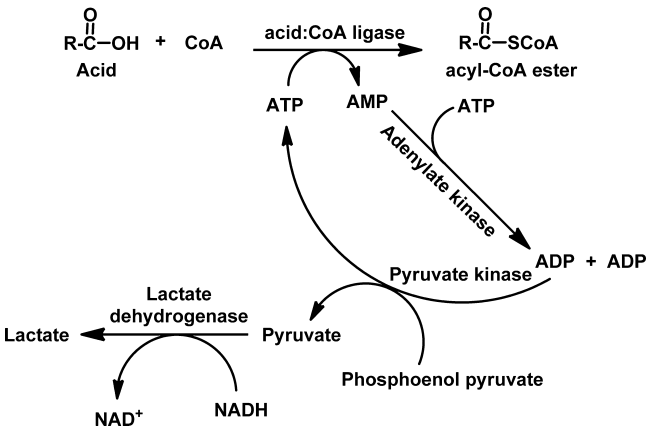
Cloning, Expression, and Protein Purification of Cinnamyl-CoA Ligase from *S. coelicolor* A3(2) (CCL). The gene encoding CCL (GI:1099823) was amplified via PCR from genomic DNA isolated from *S. coelicolor* A3(2) (ATCC) using Platinum *Pfx* DNA polymerase (Invitrogen). The PCR mixture (100 μL) contained 1 ng of plasmid DNA, 10 μL of 10× *Pfx* amplification buffer, 1 mM $MgSO_4$, dNTPs (0.4 mM each), 40 pmol of each primer (forward primer 5′-CGCGTTGGAGGACATATGTTCCGCAGCGAGTACGCAGACG-3′ and reverse primer 5′-GTGGGCGGGGATCCTCATCGCGGCTCCCTGAGCTGTCG-3′), and 5 units of Platinum *Pfx* DNA polymerase. The gene was amplified using a PTC-0200G Thermal Cycler (Bio-Rad Laboratories), with the following parameters: 94 °C for 2 min followed by 40 cycles of 94 °C for 1 min, 55 °C for 1 min and 15 s, and 68 °C for 3 min, and a final extension of 68 °C for 10 min. The amplified gene was cloned into a modified pET-15b vector (Novagen) in which the N-terminus contained 10 His residues, as previously described. The protein was expressed in *E. coli* strain Rosetta II(DE3). Transformed cells were grown at 37 °C in LB broth (supplemented with 100 μg/mL ampicillin) to an OD_{600} of 0.6, and IPTG (0.1 mM) was added to induce protein expression for 16 h. CCL was purified from the harvested cells as previously described for PCL.

Expression and Protein Purification of 4-Coumarate-CoA Ligase (Nt4CL) and Benzoate-CoA Ligase (RpBZL). Plasmids harboring the genes encoding Nt4CL (GI:12229632) and RpBZL (GI:1040685) were kindly provided by E. Pichershky.[19] The genes encoding Nt4CL and RpBZL were cloned into a pCRT7/CT-TOPO vector (Invitrogen) in which the C-terminus contained six His residues. Nt4CL was expressed in *E. coli* strain Rosetta 2 pLysS(DE3). Transformed cells were grown at 37 °C in LB broth (supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol) to an OD_{600} of 0.6, and IPTG (0.1 mM) was added to induce protein expression for 24 h at 25 °C. RpBZL was expressed in *E. coli* strain BL21(DE3). Transformed cells were grown at 37 °C in LB broth (supplemented with 100 μg/mL ampicillin) to an OD_{600} of 0.6, and IPTG (0.1 mM) was added to induce protein expression for 16 h. Nt4CL and RpBZL were purified from the harvested cells as previously described for PCL.

Cloning, Expression, and Protein Purification of Type III Polyketide Synthase from *Oryza sativa* (OsPKS). Total RNA was first extracted from 1-month-old *O. sativa* (Indica variant, a kind gift from Z. Chao Yin, Temasek Life Sciences Laboratory) plants using the RNeasy Plant Mini Kit (Qiagen). The cDNA library was constructed from total RNA using the SuperScript III First-Strand Synthesis System Kit (Invitrogen), according to the manufacturer's instructions. The gene encoding OsPKS (GI:115485731) was amplified via PCR from the *O. sativa* cDNA library using Platinum *Pfx* DNA polymerase (Invitrogen). The PCR mixture (100 μL) contained 1 ng of cDNA, 10 μL of 10× *Pfx* amplification buffer, 1 mM $MgSO_4$, dNTPs (0.4 mM each), 40 pmol of each primer (forward primer 5′-GAGAGACATATGGCAGCGGCGGTGACGGTG-3′ and reverse primer 5′-CTCTCTGGATCCTCAGGCGGCGGCGCCGGCGGTG-3′), and 5 units of Platinum *Pfx* DNA polymerase. The gene was amplified using a PTC-0200G Thermal Cycler (Bio-Rad Laboratories), with the following parameters: 94 °C for 2 min followed by 30 cycles of 94 °C for 1 min, 60 °C for 1 min and 15 s, and 68 °C for 3 min, and a final extension of 68 °C for 10 min. The amplified gene was cloned into a modified pET-15b vector (Novagen) in which the N-terminus contained 10 His residues, as previously described. The protein was expressed in *E. coli* strain Rosetta II(DE3). Transformed cells were grown at 37 °C in LB broth (supplemented with 100 μg/mL ampicillin) to an OD_{600} of 0.6, and IPTG (0.1 mM) was added to induce protein expression for 16 h. OsPKS was purified from the harvested cells as previously described for PCL.

Kinetic Assays of Acid-CoA Ligase Activity. Acid-CoA ligase activities of PCL, RtMCS, CCL, Nt4CL, and RpBZL were assayed by a continuous coupled-enzyme spectrophotometric assay, using a UV-2550 spectrophotometer (Shimadzu). The assay (200 μL at 25 °C) contained acid-CoA ligase, 100 mM Tris-HCl buffer (pH 8.0), 10 mM $MgCl_2$, 5 mM ATP, 1 mM CoA, 1.5 mM phosphoenolpyruvate, 0.16 mM NADH (340 nm; ε = 6220 M^{-1} cm^{-1}), 2 units of AK, 2 units of PK, 2 units of LDH, and 0.010−10 mM substrate. Initial rates (ν_o) were corrected for the background rate of NADH oxidation in the absence of enzyme. Kinetic parameters were determined by fitting the initial rates to the Michaelis−Menten equation using Enzfitter (Biosoft). Acid-CoA ligase activity was monitored spectrophotometrically by coupling the rate of AMP formation (ligase activity) to oxidation of NADH [using the AK−PK−LDH system (Scheme 1)]. Acid-CoA ligase activities for

Scheme 1



substrates or products (highlighted substrates in Figure S1 of the Supporting Information) that absorb at 340 nm were determined using a DTNB end-point assay measuring the amount of remaining free CoA.[21] Briefly, the assay (100 μL at 25 °C) contained acid-CoA ligase, 100 mM Tris-HCl buffer (pH 8.0), 10 mM $MgCl_2$, 5 mM ATP, 1 mM CoA, and 0.01−10 mM substrate. After incubation at 25 °C for sufficient time to convert ≤10% of the substrate to product, the reaction was stopped by addition of 900 μL of DTNB solution. The mixture was incubated at 25 °C for 5 min, and the remaining free CoA was quantitated by measuring the stoichiometrically equivalent amounts of 2-nitro-5-thiobenzoate dianion (412 nm; ε = 14150 M^{-1} cm^{-1}).

Biosynthesis of Acyl-CoA Thioesters. Acyl-CoA thioesters were prepared by reacting acid substrates and acid-CoA ligases. The reaction mixture (2 mL) contained 100 mM Tris-HCl buffer (pH 8.0), 2.5 mM ATP, 2.5 mM CoA, 5 mM $MgCl_2$, 5 mM carboxylic acid substrate, and 5 units of acid-CoA ligase. After incubation at 25 °C for 8 h, additional 2.5 mM ATP, 2.5 mM CoA, and 5 units of acid-CoA ligase were added to the mixture, and the reaction was allowed to proceed for an additional 16 h. Acid-CoA ligase was removed by centrifugation through an Amicon Ultra (10000 Da) Millipore centrifugal device. The reaction mixture was lyophilized, resuspended in 1 mL of H_2O, and loaded onto an Atlantis Prep C18 (Waters) reverse-phase high-performance liquid chromatography (HPLC) column (10 mm × 25 cm, 10 μm particle size). The column was washed with a 1:99 acetonitrile/100 mM ammonium acetate mixture (pH 6.5) at a flow rate of 5 mL/min for 4 min and eluted with a linear gradient from 1 to 70% acetonitrile [from 99 to 30% 100 mM ammonium acetate (pH 6.5), respectively] over 60 min. Fractions containing the acyl-CoA thioester (monitored at 230, 240, and 259 nm) were pooled and lyophilized. The identity of the acyl-CoA thioester was determined by high-resolution mass spectrometry performed at the Mass Spectrometry Laboratory at the Department of Chemistry, National University of Singapore.

Construction of RtMCS−Nt4CL and RtMCS−RpBZL Chimeras. With the expectation that RtMCS, Nt4CL, and RpBZL are enzymatically and structurally homologous to other adenylate-forming acid-CoA ligases, we identified the specificity loops that form the acyl-binding pockets from a sequence alignment based upon a superposition of the structures of MatB [Protein Data Bank (PDB) entry 3NYQ], Pt4CL (4-coumarate:CoA ligase from *Populos tomentosa*, PDB entry 3NI2), and CBL (PDB entry 3CW9), respectively;[18,22] these loops, highlighted in green in Figure 1, are represented by the sequences boxed in magenta in Figure S2 of the Supporting Information. RtMCS−Nt4CL and RtMCS−RpBZL chimeras were constructed by PCR-based loop swapping, with the primers for construction of the chimeras detailed in the Supporting Information. RtMCS−Nt4CL and RtMCS−RpBZL chimeras were RtMCS mutants that had their loops forming the acyl-binding pockets replaced with the corresponding loops from Nt4CL and RpBZL, respectively (exact sequences are highlighted by the magenta box in Figure S2 of the Supporting Information). These chimeras were verified by DNA sequencing and purified as previously described for the wild-type protein.

Biosynthesis of Polyketides. Polyketides were biosynthesized in vitro from precursor acid substrates using the purified acid-CoA ligases Nt4CL and RtMCS, respectively, and the Type III PKS OsPKS. Initial biosynthesis of the starter CoA thioester was conducted in a 1 mL reaction mixture containing 100 mM Tris-HCl (pH 8.0), 2 mM ATP, 2 mM CoA, 5 mM $MgCl_2$, 2 mM carboxylic acid, and 5 units of Nt4CL. After incubation at 25 °C for 24 h, additional extender acid precursor, cofactors, and enzymes were added to a final reaction volume of 2 mL, containing 100 mM Tris-HCl (pH 8.0), 5 mM ATP, 0.5 mM CoA, 5 mM $MgCl_2$, ≤1 mM starter CoA, 5 mM malonic acid (extender acid precursor), 5 units of Nt4CL, 5 units of RtMCS, and 0.5 mg/mL OsPKS. The reaction mixture was further incubated at 25 °C for an additional 24 h. The enzymes were removed by centrifugation through an Amicon Ultra (10000 Da) Millipore centrifugal device. The reaction mixture was lyophilized, resuspended in 1 mL of H_2O, and loaded onto an Atlantis Prep C18 (Waters) reverse-phase HPLC column (10 mm × 25 cm, 10 μm particle size). The column was washed with a 1:99 acetonitrile/100 mM ammonium acetate mixture (pH 6.5) at a flow rate of 5 mL/min for 4 min and eluted with a linear gradient from 1 to 70% acetonitrile [99 to 30% 100 mM ammonium acetate (pH 6.5), respectively] for 20 min, followed by a step elution with a 70:30 acetonitrile/100 mM ammonium acetate mixturre (pH 6.5) for 4 min. Fractions containing the polyketide (monitored at 230, 240, and 259 nm) were pooled and lyophilized. The identity of the polyketide was determined by high-resolution mass spectrometry as previously described.

RESULTS AND DISCUSSION

The biosynthesis of "unnatural" polyketides has long been the goal of synthetic enzymologists since the discovery of

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568−4579

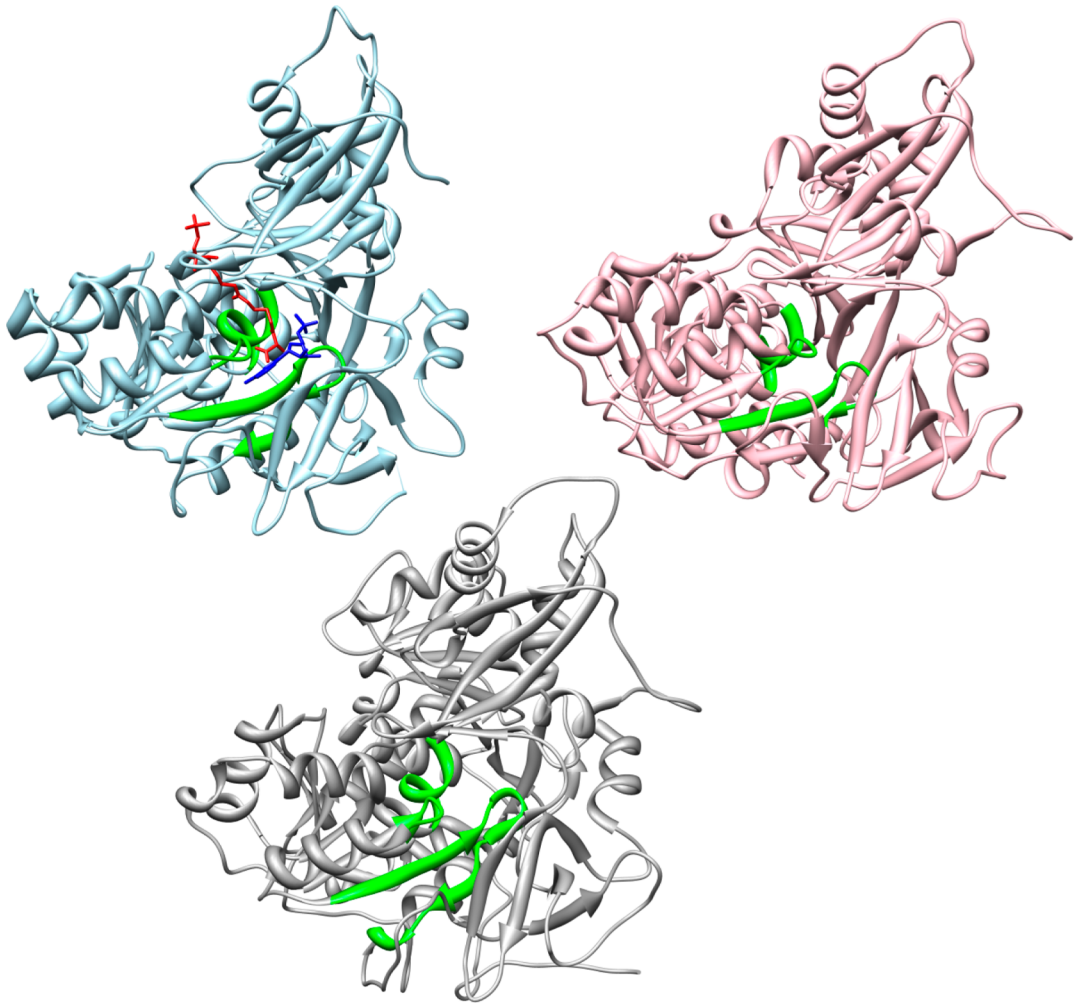


Figure 1. Ribbon representations of the structures of MatB (light blue), CBL (pink), and Pt4CL (gray). The substrate specificity loops, corresponding to the sequences boxed in magenta in Figure S2 of the Supporting Information, are highlighted in green. The CoA and AMP ligands bound in the active site of MatB are colored red and blue, respectively. This figure was produced using the UCSF Chimera package, from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, CA (supported by National Institutes of Health Grant P41 RR-01081).[29]

polyketide synthases. Precursor-directed biosyntheses, coupled with combinatorial efforts, have been successfully demonstrated in vitro[23] and in vivo.[24] However, to fully realize the potential of synthetic polyketide biosynthesis, executed through the construction of de novo polyketide biosynthetic pathways in cells, the major challenge of elucidating the substrate specificity and function of the enzymes (acid-CoA ligases and PKSs) involved has to be met.[5,25,26] This study focuses on the assembly of a toolkit that catalogs a list of acyl-CoA thioesters that can be biosynthesized by acid-CoA ligases, defining the substrate specificities of these enzymes for subsequent polyketide production. To assemble such a list that provides enzymatic routes to the biosynthesis of novel polyketides (despite the mechanistic diversity among the three paradigmatic classes of PKSs, polyketide biosynthesis is still dependent on the availability of novel acyl-CoA precursors and extenders), we sought to identify five acid-CoA ligases that would catalyze the formation of thioesters between CoA and a "chemically diverse" set of carboxylic acids.

Assembling a Set of Acid-CoA Ligases for Acyl-CoA Library Biosynthesis. Five acid-CoA ligases, namely, RtMCS, CCL, Nt4CL, RpBZL, and PCL, were selected on the basis of the availability of genomic DNA and their reported and/or expected substrate specificities. We proceeded to determine the extent of substrate promiscuity exhibited by these enzymes by screening each of them against 123 carboxylic acids for acid-CoA ligase activity. We identified carboxylic acid substrates for the respective acid-CoA ligases using spectrophotometric assays that detected either the conversion of ATP to AMP [for the enzyme-coupled assay (Scheme 1)] or the consumption of CoA (for the DTNB assay).

We sought to corroborate the results of our enzymatic assays by analyzing the HPLC chromatograms of the reaction mixtures of the acid-CoA ligases. We assessed the production of the respective acyl-CoA thioester products (identified from our enzymatic assays) by detecting "new" acyl-CoA thioester peaks present within the in vitro biosynthesis reaction mixtures. A representative analytical HPLC chromatogram depicting the presence of the acyl-CoA thioester product is shown in Figure 2. We sought to confirm the substrate profiles of the respective acid-CoA ligases by determining the mass of the acyl-CoA thioester products using mass spectrometry. The masses of all of the thioester products identified in our study matched their corresponding expected masses. Altogether, we have exper-

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568−4579

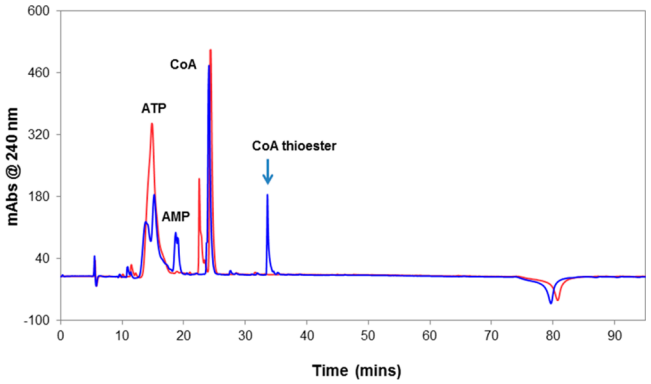


Figure 2. Analytical HPLC chromatogram showing the biosynthesis of 4-hydroxycinnamyl-CoA. HPLC chromatogram overlays of the nonenzymatic control (red) and reaction mixture (blue) containing 100 mM Tris-HCl buffer (pH 8.0), 5 mM ATP, 5 mM CoA, 5 mM $MgCl_2$, 5 mM 4-hydroxycinnamic acid substrate, and 5 units of CCL. The 4-hydroxycinnamyl-CoA product peak (labeled as CoA thioester) is denoted with a blue arrow.

imental evidence of the biosynthesis of a library of 79 acyl-CoA thioesters, including two previously unreported malonyl-CoA derivatives (3-thiophenemalonyl-CoA and phenylmalonyl-CoA).

Biosynthesis of a Library of Acyl-CoA Thioesters. For the sake of clarity and easy reference, we grouped the library of acyl-CoA thioesters according to their chemical classes and listed the kinetic parameters of the corresponding acid-CoA ligases in the tables. We report the kinetic parameters for the biosynthesis of derivatives of cinnamyl-CoA and 3-phenylpropanoyl-CoA by CCL and Nt4CL, respectively, in Table 1, derivatives of benzoyl-CoA by BZL in Table 2, derivatives of phenylacetyl-CoA by PCL in Table 3, saturated aliphatic CoA thioesters by CCL, Nt4CL, BZL, and PCL in Table 4, unsaturated aliphatic CoA thioesters by BZL and PCL in Table 5, bicyclic aromatic CoA thioesters by Nt4CL in Table 6, and derivatives of malonyl-CoA by RtMCS and PCL in Table 7.

In the exploration of the substrate promiscuities of the acid-CoA ligases, we noticed trends that may aid in the selection of additional substrates and, hence, in the future expansion of the toolkit for polyketide biosynthesis. Nt4CL exhibited a preference for cinnamate (and its derivatives) substrates over 3-phenylpropanoates (and its derivatives): the presence of the unsaturated carbon−carbon double bond in cinnamates corresponded to a decrease in K_M (during thioester bond formation) by 1 order of magnitude (10-fold), when compared to those of the phenylpropanoates. The corresponding decreases in K_M between derivatives of cinnamic acid and 3-phenylpropanoic acid are listed in Table 1. Derivatization of the cinnamic acid substrate by substitutions at the ortho, meta, and para positions of the aromatic ring affected the catalytic efficiency (k_{cat}/K_M) of the Nt4CL-catalyzed reaction in the following manner: substitutions at the meta position had the

Table 1. Kinetic Parameters for the Biosynthesis of Derivatives of Cinnamyl-CoA and 3-Phenylpropanoyl-CoA[a]

	CCL			Nt4CL		
substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
cinnamic acid	$\leq 2.8 \times 10^{-5}$			0.32 ± 0.015	0.14 ± 0.028	2.2×10^3
2-fluorocinnamic acid	ND	ND	ND	$(9.1 \pm 1.0) \times 10^{-2}$	0.14 ± 0.089	640
2-hydroxycinnamic acid	ND	ND	ND	$(9.8 \pm 0.87) \times 10^{-2}$	0.12 ± 0.055	790
3-chlorocinnamic acid	ND	ND	ND	$(9.6 \pm 0.50) \times 10^{-2}$	$(4.9 \pm 1.4) \times 10^{-2}$	2.0×10^3
3-fluorocinnamic acid	ND	ND	ND	0.21 ± 0.034	$(3.1 \pm 0.99) \times 10^{-2}$	6.9×10^3
3-chloro-4-methoxycinnamic acid	ND	ND	ND	0.16 ± 0.0067	$(2.2 \pm 0.56) \times 10^{-2}$	7.3×10^3
3-methoxy-4-hydroxycinnamic acid	ND	ND	ND	0.14 ± 0.013	$(7.7 \pm 3.5) \times 10^{-2}$	1.8×10^3
4-chlorocinnamic acid	ND	ND	ND	$(7.6 \pm 1.0) \times 10^{-2}$	0.14 ± 0.088	560
4-fluorocinnamic acid	$(6.8 \pm 0.26) \times 10^{-3}$	$(2.7 \pm 0.51) \times 10^{-2}$	250	0.11 ± 0.0050	$(3.1 \pm 0.99) \times 10^{-2}$	3.5×10^3
4-hydroxycinnamic acid	$(1.3 \pm 0.052) \times 10^{-2}$	2.1 ± 0.20	6.0	0.19 ± 0.014	$(4.0 \pm 1.6) \times 10^{-2}$	4.8×10^3
4-methoxycinnamic acid	$\leq 1.6 \times 10^{-4}$			$(6.4 \pm 0.17) \times 10^{-2}$	$(4.4 \pm 1.0) \times 10^{-3}$	1.4×10^4
4-methylcinnamic acid	$\leq 3.4 \times 10^{-4}$			$(9.5 \pm 1.1) \times 10^{-2}$	$(6.0 \pm 3.6) \times 10^{-2}$	1.6×10^3
α-fluorocinnamic acid	ND	ND	ND	$(5.6 \pm 0.28) \times 10^{-3}$	$(8.0 \pm 1.8) \times 10^{-2}$	70
α-methylcinnamic acid	*b*	*b*	0.74	0.16 ± 0.014	0.54 ± 0.11	310
3-phenylpropanoic acid	*b*	*b*	0.63	0.35 ± 0.034	1.8 ± 0.57	190
3-(3-chlorophenyl)propanoic acid	*b*	*b*	0.29	$(6.9 \pm 0.71) \times 10^{-2}$	0.26 ± 0.085	270
3-(3-methoxyphenyl)propanoic acid	*b*	*b*	0.32	0.11 ± 0.0084	0.11 ± 0.037	990
3-(3-chloro-4-methoxyphenyl) propanoic acid	$\leq 1.2 \times 10^{-2}$			$(2.1 \pm 0.23) \times 10^{-3}$	0.39 ± 0.10	5.4
3-(3,4-dihydroxyphenyl) propanoic acid	ND	ND	ND	0.12 ± 0.0096	$(3.0 \pm 0.86) \times 10^{-2}$	4.2×10^3
3-(3,5-difluorophenyl) propanoic acid	$(1.6 \pm 0.53) \times 10^{-2}$	25 ± 13	0.64	ND	ND	ND
3-(4-fluorophenyl)propanoic acid	*b*	*b*	0.32	$(4.1 \pm 0.039) \times 10^{-3}$	0.23 ± 0.0076	18
3-(4-methoxyphenyl)propanoic acid	*b*	*b*	0.19	$(8.4 \pm 1.9) \times 10^{-3}$	0.11 ± 0.037	11

[a]ND, no detectable activity. [b]Saturation kinetics could not be attained.

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568−4579

Table 2. Kinetic Parameters for the Biosynthesis of Derivatives of Benzoyl-CoA

substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
benzoic acid	8.4 ± 0.29	(1.2 ± 0.30) × 10^{-3}	7.3 × 10^6
2-bromobenzoic acid	(1.4 ± 0.055) × 10^{-2}	(9.0 ± 1.6) × 10^{-2}	150
2-chlorobenzoic acid	0.33 ± 0.012	0.14 ± 0.020	2.3 × 10^3
2-fluorobenzoic acid	10 ± 0.46	(1.3 ± 0.23) × 10^{-2}	7.4 × 10^5
2-hydroxybenzoic acid	0.18 ± 0.0022	(1.2 ± 0.091) × 10^{-2}	1.5 × 10^4
2-iodobenzoic acid	≤9.1 × 10^{-5}		
2-methoxybenzoic acid	0.10 ± 0.0096	1.2 ± 0.41	86
2-methylbenzoic acid	0.42 ± 0.047	0.15 ± 0.026	2.9 × 10^3
2,3-dihydroxybenzoic acid	(4.5 ± 0.27) × 10^{-2}	1.1 ± 0.19	42
2,4-dihydroxybenzoic acid	(1.9 ± 0.17) × 10^{-2}	0.38 ± 0.12	50
2,5-dihydroxybenzoic acid	*a*	*a*	2.2
2,6-difluorobenzoic acid	7.5 ± 1.3	1.6 ± 0.42	4.9 × 10^3
3-aminobenzoic acid	1.3 ± 0.090	0.95 ± 0.14	1.4 × 10^3
3-bromobenzoic acid	*a*	*a*	1.7
3-chlorobenzoic acid	0.43 ± 0.033	4.7 ± 0.49	90
3-fluorobenzoic acid	7.3 ± 0.59	(6.1 ± 1.8) × 10^{-2}	1.2 × 10^5
4-aminobenzoic acid	(2.1 ± 0.023) × 10^{-2}	(9.3 ± 0.38) × 10^{-2}	230
4-bromobenzoic acid	(1.9 ± 0.37) × 10^{-3}	0.52 ± 0.30	3.6
4-chlorobenzoic acid	0.23 ± 0.025	0.74 ± 0.21	310
4-fluorobenzoic acid	6.5 ± 0.27	(1.1 ± 0.15) × 10^{-2}	5.9 × 10^5

[a]Saturation kinetics could not be attained.

Table 3. Kinetic Parameters for the Biosynthesis of Derivatives of Phenylacetyl-CoA

substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
phenylacetic acid	9.4 ± 0.21	(1.6 ± 0.17) × 10^{-2}	6.0 × 10^5
phenylpyruvic acid	(4.1 ± 0.62) × 10^{-2}	7.1 ± 1.4	5.7
2-hydroxyphenylacetic acid	2.0 ± 0.069	0.26 ± 0.034	7.8 × 10^3
4-hydroxyphenylacetic acid	4.9 ± 0.56	4.1 ± 0.83	1.2 × 10^3
4-methoxyphenylacetic acid	0.25 ± 0.031	3.0 ± 0.55	86
phenoxyacetic acid	(5.9 ± 0.57) × 10^{-2}	0.26 ± 0.098	220

Table 4. Kinetic Parameters for the Biosynthesis of Saturated Aliphatic CoA Thioesters[a]

	CCL			Nt4CL		
substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
propanoic acid	ND	ND	ND	ND	ND	ND
butanoic acid	*b*	*b*	0.37	ND	ND	ND
pentanoic acid	*b*	*b*	2.5	ND	ND	ND
hexanoic acid	(5.6 ± 0.20) × 10^{-2}	(9.0 ± 1.0) × 10^{-2}	620	(2.5 ± 0.12) × 10^{-2}	(4.9 ± 0.91) × 10^{-2}	510
heptanoic acid	ND	ND	ND	(6.3 ± 0.20) × 10^{-2}	(3.4 ± 0.50) × 10^{-2}	1.8 × 10^3
octanoic acid	ND	ND	ND	(7.5 ± 0.34) × 10^{-2}	(4.4 ± 0.93) × 10^{-2}	1.7 × 10^3
nonanoic acid	ND	ND	ND	0.17 ± 0.012	(2.4 ± 0.55) × 10^{-2}	7.2 × 10^3
decanoic acid	ND	ND	ND	(6.6 ± 0.15) × 10^{-2}	(4.1 ± 0.44) × 10^{-2}	1.6 × 10^3
	BZL			PCL		
substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
propanoic acid	*b*	*b*	15	ND	ND	ND
butanoic acid	*b*	*b*	66	(6.8 ± 1.5) × 10^{-2}	9.9 ± 3.8	6.9
pentanoic acid	*b*	*b*	7.0	1.3 ± 0.15	12 ± 1.8	110
hexanoic acid	≤2.9 × 10^{-4}			0.67 ± 0.13	6.1 ± 1.9	110
heptanoic acid	ND	ND	ND	0.21 ± 0.028	5.1 ± 1.2	40
octanoic acid	ND	ND	ND	(3.5 ± 0.34) × 10^{-2}	1.0 ± 0.30	34
nonanoic acid	ND	ND	ND	(1.7 ± 0.14) × 10^{-2}	3.7 ± 0.82	4.6
decanoic acid	ND	ND	ND	ND	ND	ND

[a]ND, no detectable activity. [b]Saturation kinetics could not be attained.

least effect on catalytic efficiency, followed by substitutions at the para-position, while substitutions at the ortho position significantly decreased catalytic efficiency [comparative catalytic efficiencies: 2-fluorocinnamate < 4-fluorocinnamate < 3-fluorocinnamate (Table 1)]. Similar effects of positional substitutions of the aromatic ring on catalytic efficiencies of Nt4CL on the 3-phenylpropanoates (and its derivatives) were observed.

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568−4579

Table 5. Kinetic Parameters for the Biosynthesis of Unsaturated Aliphatic CoA Thioesters[a]

	BZL			PCL		
substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
2-butenoic acid	2.4 ± 0.33	6.2 ± 1.3	390	ND	ND	ND
3-butenoic acid	1.3 ± 0.11	22 ± 2.6	60	ND	ND	ND
2-methyl-2-butenoic acid	0.34 ± 0.015	(3.5 ± 0.76) × 10^{-3}	9.8 × 10^4	(5.9 ± 0.67) × 10^{-3}	1.9 ± 0.54	3.1
3-methyl-2-butenoic acid	2.8 ± 0.20	1.7 ± 0.32	1.6 × 10^3	ND	ND	ND
2-pentenoic acid	3.6 ± 0.26	1.1 ± 0.25	3.3 × 10^3	*b*	*b*	19
3-pentenoic acid	*b*	*b*	34	6.7 ± 0.51	6.0 ± 0.74	1.1 × 10^3
4-pentenoic acid	5.3 ± 0.75	0.69 ± 0.27	7.6 × 10^3	1.2 ± 0.18	15 ± 2.8	78
3-methyl-4-pentenoic acid	ND	ND	ND	0.16 ± 0.063	30 ± 14	5.5
3-hexenoic acid	ND	ND	ND	7.4 ± 0.30	5.2 ± 0.35	1.4 × 10^3
5-hexenoic acid	(5.5 ± 0.42) × 10^{-3}	7.1 ± 1.1	0.78	3.9 ± 0.63	5.2 ± 1.4	740

[a]ND, no detectable activity. [b]Saturation kinetics could not be attained.

Table 6. Kinetic Parameters for the Biosynthesis of Bicyclic Aromatic CoA Thioesters (Nt4CL)

substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
1-naphthalenecarboxylic acid	(1.4 ± 0.093) × 10^{-2}	0.25 ± 0.052	56
2-naphthalenecarboxylic acid	(3.0 ± 0.20) × 10^{-2}	(3.5 ± 1.1) × 10^{-2}	860
2-quinolinecarboxylic acid	(3.0 ± 0.34) × 10^{-3}	0.20 ± 0.070	15
3-quinolinecarboxylic acid	(2.7 ± 0.19) × 10^{-2}	0.61 ± 0.13	44

Table 7. Kinetic Parameters for the Biosynthesis of Derivatives of Malonyl-CoA[a]

	RtMCS			PCL		
substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
malonic acid	86 ± 2.3	3.7 ± 0.19	2.3 × 10^4	ND	ND	ND
methylmalonic acid	9.2 ± 0.90	3.5 ± 0.83	2.7 × 10^3	ND	ND	ND
ethylmalonic acid	2.0 ± 1.2	57 ± 38	35	ND	ND	ND
isopropylmalonic acid	23 ± 8.9	47 ± 22	490	ND	ND	ND
butylmalonic acid	(2.5 ± 0.52) × 10^{-2}	29 ± 7.8	0.86	ND	ND	ND
allylmalonic acid	1.6 ± 0.25	11 ± 2.3	140	ND	ND	ND
hydroxymalonic acid	0.49 ± 0.12	30 ± 9.4	16	ND	ND	ND
3-thiophenemalonic acid	ND	ND	ND	4.9 ± 0.60	1.1 ± 0.40	4.5 × 10^3
phenylmalonic acid	ND	ND	ND	5.0 ± 0.086	0.27 ± 0.017	1.8 × 10^4

[a]ND, no detectable activity.

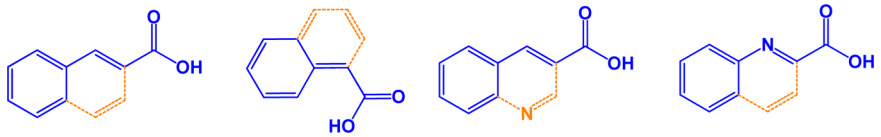


Figure 3. Structural mimicry of bicyclic aromatic acid substrates of Nt4CL. Chemical structures of 1-naphthalenecarboxylic acid, 2-naphthalenecarboxylic acid, 2-quinolinecarboxylic acid, and 3-quinolinecarboxylic acid (from left to right, respectively). The cinnamic acid mimicry is highlighted in blue.

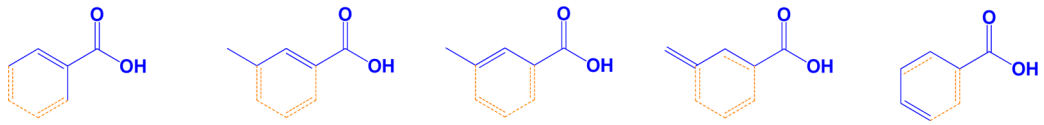


Figure 4. Structural mimicry of unsaturated aliphatic acid substrates of BZL. Chemical structures of 2-methyl-2-butenoic acid, 2-pentenoic acid, 3-pentenoic acid, 4-pentenoic acid, and 5-hexenoic acid (from left to right, respectively) highlighted in blue. The benzoic acid mimicry is denoted with broken orange lines.

Uncovering Unexpected Substrate Promiscuities. Apart from derivatives of the cognate substrates, we found that some of the acid-CoA ligases could accept "chemically distinct" acid substrates beyond their canonical substrate pools. Nt4CL was able to catalyze the formation of a thioester between bicyclic aromatic acids (such as 1-naphthalenecarboxylic acid) and CoA. We attribute this promiscuous substrate range to "substrate mimicry", a theme that could generally account for the range of substrate promiscuities reported in this study. Because the cognate substrate for Nt4CL is cinnamic acid, acids that bear a structural resemblance to the substrate cinnamic acid could be potential substrates for acid-CoA

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568−4579

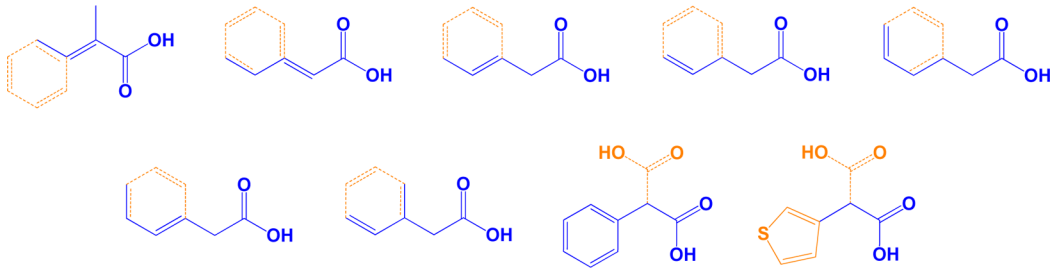


Figure 5. Structural mimicry of unsaturated aliphatic acid substrates and malonic acid derivatives of PCL. Chemical structures of 2-methyl-2-butenoic acid, 2-pentenoic acid, 3-pentenoic acid, 4-pentenoic acid, 5-hexenoic acid, 3-hexenoic acid, 3-methyl-4-pentenoic acid (highlighted in solid blue lines), phenylmalonic acid, and 3-thiophenemalonic acid (from left to right and from top to bottom, respectively). The phenylacetic acid mimicry is denoted by noncontinuous orange lines, with the exception of phenylmalonic acid and 3-thiophenemalonic acid.

thioesterification (Figure 3 illustrates the bicyclic aromatic acid substrates of Nt4CL and the detailing of the structural mimicry contained within the corresponding substrates). However, not all bicyclic aromatic acids are good structural mimics of the cinnamic acid template: because of the positioning of the carboxylate group that prevented the structural mimicry, Nt4CL did not accept 4-quinolinecarboxylic acid and isoquinoline-1-carboxylic acid as substrates (chemical structures are shown in Figure S1 of the Supporting Information).

Apart from a wide range of benzoic acid derivatives (Table 2), RpBZL could biosynthesize unsaturated alkyl-CoAs [such as *trans*-2-butenoyl-CoA or crotonoyl-CoA (Table 5)], providing novel enzymatic routes for polyketide biosynthesis efforts. Again, we attribute this substrate promiscuity to "structural mimicry": RpBZL could utilize unsaturated aliphatic acids that bore a structural resemblance to the cognate benzoic acid substrate (Figure 4 illustrates the unsaturated aliphatic acid substrates of RpBZL and the corresponding structural mimicry). The phenomenon of catalyzing promiscuous substrates that bore a resemblance to the cognate substrate was also observed with PCL: PCL accepted derivatives of phenylacetic acids, unsaturated aliphatic acids, and two malonic acid derivatives (3-thiophenemalonic acid and phenylmalonic acid) as substrates for thioesterification to CoA (Figure 5).

Our study of RtMCS was consistent with previous findings that indicated the expected substrate promiscuity of the acid-CoA ligase toward a range of malonic acid derivatives (Table 7). However, we noticed that RtMCS could not catalyze thioesterification of 3-thiophenemalonic acid and phenylmalonic acid to CoA. Instead, after a serendipitous search using other acid-CoA ligases, we report the biosynthesis of two novel malonyl-CoA derivatives, 3-thiophenemalonyl-CoA and phenylmalonyl-CoA, using PCL as the acid-CoA ligase. To the best of our knowledge, this is the first description of the biosynthesis of these malonyl-CoA derivatives that we hope would aid in future polyketide biosynthetic efforts (by all three paradigmatic classes of PKSs).

We observed that structural mimicry or steric effects did not solely account for substrate specificity or promiscuity of acid-CoA ligases. The polarity of the carboxylic acid substrate contributed, to a certain extent, to the specificity of the acid-CoA ligase. We observed that RpBZL could not utilize pyridinecarboxylic acids as substrates: presumably, the presence of the N atom in the aromatic ring would have changed the polarity of the substrate! As such, RpBZL could not catalyze thioesterification between pyridine-3-carboxylic acid, pyridine-2-carboxylic acid, pyridinecarboxylic acid, 2-chloropyridine-3-carboxylic acid, and pyridine-4-carboxylic acid and CoA (the catalytic efficiencies of thioesterification of the corresponding substrates benzoic acid and 2-chlorobenzoic acid are 7.3×10^6 and 2.3×10^3 M^{-1} s^{-1}, respectively). Analogous to the polarity effect observed in RpBZL substrate specificity, Nt4CL was unable to utilize carboxylic acids with polarities distinct from that of the cinnamic acid template: although 2-quinoxaline-carboxylic acid and benzylmalonic acid bear a structural resemblance to the cinnamic acid template, their differences in polarity prevented subsequent thioesterification to the CoA moiety.

Rational Design of Promiscuous Acid-CoA Ligases. In the process of assembling a toolkit to provide the means of delivering novel acyl-CoA precursors necessary for polyketide biosynthesis, we noticed the differing solubilities and overall protein yields of the five acid-CoA ligases: the protein yields (per liter of culture) for RtMCS, RpBZL, PCL, CCL, and Nt4CL were 80, 10, 12, 2.5, and 2 mg, respectively. Because the grafting of active site loops has been successfully employed to effect changes in substrate specificity,[27,28] we sought to construct chimeric acid-CoA ligases with relatively high yields and solubilities (>50 mg/L of culture) and a broader substrate range so that these enzymes can be used for polyketide biosynthesis. Using a structural superposition of orthologous acid-CoA ligases MatB (for RtMCS), 4-coumarate-CoA ligase from *P. tomentosa* (for Nt4CL), and CBL (for RpBZL), we identified the specificity loops for the construction of chimeras (Figure 1). With the expectation that RtMCS, Nt4CL, and RpBZL are structurally homologous, we chose to graft the specificity loops from RpBZL and Nt4CL onto the RtMCS template with the intention of constructing chimeric acid-CoA ligases (RtMCS−RpBZL and RtMCS−Nt4CL chimeras, respectively) with yields and solubilities comparable to those of RtMCS, along with a broader substrate range beyond malonic acid and its derivatives.

The origin of substrate specificities of acid-CoA ligases has been traced to the substrate-binding pockets of these enzymes;[10,12,13] residues within the pocket are part of specificity loops that define the overall substrate-determining architecture. In a simplistic attempt to "transplant" substrate specificities between acid-CoA ligases, we hypothesized that the architecture of the substrate-binding pocket resided within the residues that comprise the specificity loops. Figure 6 illustrates the architecture of the substrate-binding pocket of MatB as defined by the specificity loops (highlighted in green).

The acid-CoA ligase RtMCS−RpBZL and RtMCS−Nt4CL chimeras were heterologously expressed with yields (per liter of culture) of 27 and 6 mg, respectively. In comparison with the yields of wild-type RtMCS (80 mg/L of culture), RpBZL (10

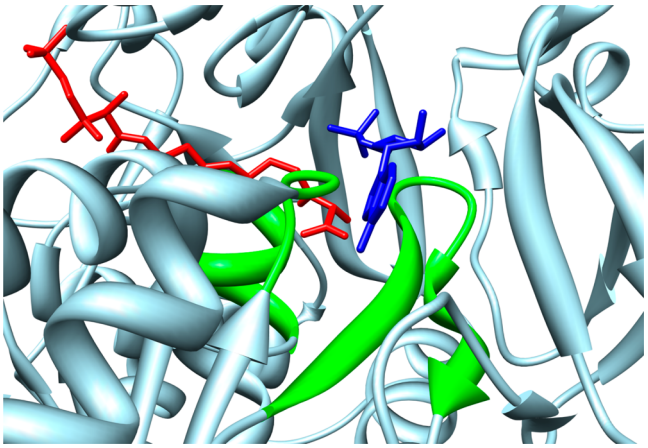
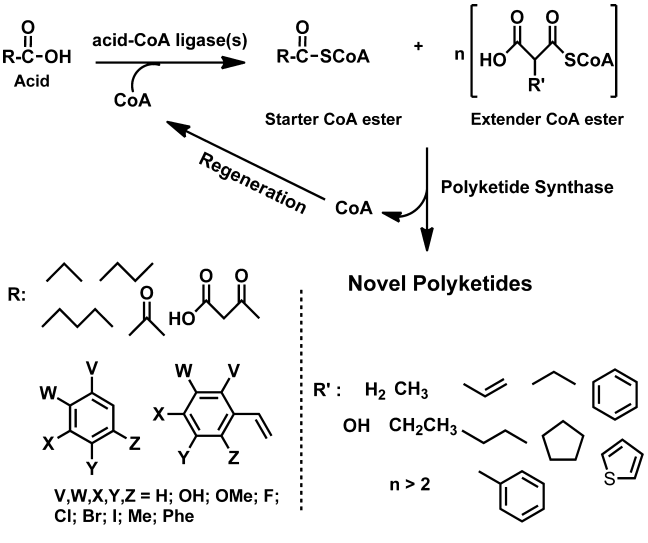


Figure 6. Substrate-binding pocket of MatB. Ribbon representation of the structure of MatB (light blue), showing the active site and the substrate-binding pocket defined by substrate specificity loops (highlighted in green), corresponding to the sequences boxed in magenta in Figure S2 of the Supporting Information. The CoA and AMP ligands bound in the active site of MatB are colored red and blue, respectively. This figure was produced using the UCSF Chimera package.

mg/L of culture), and Nt4CL (2 mg/L of culture), the chimeric mutants were less soluble than the RtMCS template but were correspondingly more soluble than RpBZL (an increase of 2.7-fold in protein yield per liter of culture) and Nt4CL (an increase of 3-fold in protein yield per liter of culture). However, we did not detect any acid-CoA ligase activity with the RtMCS–RpBZL or RtMCS–Nt4CL chimera against the 123 carboxylic acids used in the study. Thus, although orthologous design efforts through site-directed mutagenesis and loop grafting have been successfully attempted in the past, the principles underlying the introduction of new substrate specificities to old scaffolds still remain unknown.

Implications for Biosynthesis of Polyketides. We envision that the availability of a library of acyl-CoA thioesters (or knowledge of the enzymatic route of biosynthesis) would facilitate downstream biosynthesis of novel polyketides. As illustrated in Scheme 2, the generation of novel polyketides

Scheme 2



using precursor-directed combinatorial biosynthesis can be achieved both in vitro and in vivo. As mentioned above, this concept of polyketide biosynthesis is not new; rather, the novelty (of this study) lies in the establishment of suitable route(s) of biosynthesis for novel starter and extender CoA thioesters. Our study provides a means (toolkit) of accessing an ever expanding lexicon of CoA thioesters required for the biosynthesis of novel biomedically important polyketide-based therapeutics. Using a previously unreported Type III PKS from *O. sativa* (OsPKS), we demonstrated the utility of our toolkit through the biosynthesis of unnatural fluorinated polyketides (Figure 7). We observed the formation of these polyketides

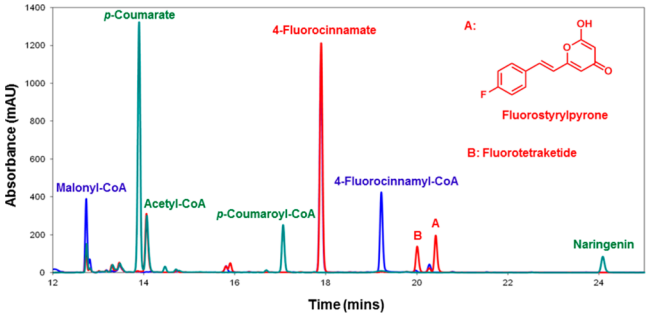


Figure 7. Analytical HPLC chromatogram showing the biosynthesis of unnatural polyketides. HPLC chromatogram overlays of the non-OsPKS control (blue) and reaction mixture (red) containing 100 mM Tris-HCl buffer (pH 8.0), 5 mM ATP, 0.5 mM CoA, 5 mM $MgCl_2$, 1 mM 4-fluorocinnamic acid (starter acid precursor), 5 mM malonic acid (extender acid precursor), 5 units of Nt4CL, 5 units of RtMCS, and 0.5 mg/mL OsPKS. The HPLC chromatogram overlay of the reaction mixture (green) containing *p*-coumaric acid (4-hydroxycinnamic acid), instead of 4-fluorocinnamic acid, shows the presence of the biosynthesized narigenin polyketide product. OsPKS will biosynthesize naringenin using *p*-coumaroyl-CoA and malonyl-CoA as the starter CoA and extender CoA, respectively; OsPKS will correspondingly biosynthesize the triketide fluorostyrylpyrone and the fluorotetraketide(s) as polyketide products using 4-fluorocinnamyl-CoA and malonyl-CoA as the starter CoA and extender CoA, respectively.

using 4-fluorocinnamate and malonate as the precursor acid substrates (generating 4-fluorocinnamyl-CoA as the starter CoA and malonyl-CoA as the extender CoA, by Nt4CL and RtMCS, respectively); the presence of the unnatural triketide fluorostyrylpyrone (compound **A** in Figure 7) was confirmed by high-resolution mass spectrometry (theoretical mass of 232.0614, experimental mass of 232.0612 ± 0.0002). The presence of the fluorotetraketide product(s) was similarly determined by mass spectrometry (experimental mass of 273.86); because of the identical mass of the possible cyclized products (via C6 → C1 Claisen-type condensation, C2 → C7 aldol condensation, or C5-O → C1 lactonization), we could not determine the chemical structure of the biosynthesized fluorotetraketide (a scheme detailing the possible cyclization routes during OsPKS-catalyzed polyketide biosynthesis is represented in Figure S4 of the Supporting Information). We are currently in the process of using this toolkit to develop polyketide-based therapeutics with antiaging, anticancer, and antimicrobial properties.

Conclusions. We have determined the extent of substrate promiscuity of five acid-CoA ligases, namely, RtMCS, RpBZL, PCL, CCL, and Nt4CL, by using a library of 123 carboxylic acids that include malonates, cinnamates, 3-phenylpropanoates, phenylacetates, benzoates, saturated and unsaturated aliphatics,

dx.doi.org/10.1021/bi300425j | *Biochemistry* 2012, 51, 4568–4579

naphthalene-, pyridine-, and quinoline-carboxylates, and their derivatives. We report a continuous coupled enzyme spectrophotometric assay for the routine determination of acid-CoA ligase activity. By exploring substrate promiscuities across different carboxylic acid classes beyond that of the cognate substrates, we discovered new enzymatic routes for the biosynthesis of previously unreported CoA thioesters, including two novel malonyl-CoA derivatives, 3-thiophenemalonyl-CoA and phenylmalonyl-CoA. This study illustrates the utility and importance of determining substrate promiscuities beyond conventional substrate pools and provides an additional tool(s) for establishing precursor-directed combinatorial polyketide biosynthesis.

■ ASSOCIATED CONTENT

Supporting Information

Chemical structures of the 123 carboxylic acids used in the study (Figure S1), sequence alignment of the acid-CoA ligases (Figure S2), sequences of primers used in the study (Table S3), and the cyclization scheme for fluorinated polyketide biosynthesis by OsPKS (Figure S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Dr., Singapore 117597. Phone: (+65) 6516-8624. Fax: (+65) 6779-1453. E-mail: bchyws@nus.edu.sg.

Funding

This research was supported by grants from the Ministry of Education of Singapore.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Professor Eran Pichershky for the kind gifts of Nt4CL and RpBZL.

■ ABBREVIATIONS

CoA, coenzyme A; PKS, polyketide synthase; RtMCS, acid-CoA ligase from *R. trifolii*; MatB, malonyl-CoA synthetase from *S. coelicolor*; CCL, cinnamyl-CoA ligase from *S. coelicolor*; Nt4CL, 4-coumarate-CoA ligase from *N. tabacum*; RpBZL, benzoate-CoA ligase from *Rh. palustri*; PCL, phenylacetate-CoA ligase from *S. coelicolor*; Pt4CL, 4-coumarate:CoA ligase from *P. tomentosa*; CBL, 4-chlorobenzoate:CoA ligase; OsPKS, Type III polyketide synthase from *O. sativa*; ATCC, American Type Culture Collection; AK, adenylate kinase; PK, pyruvate kinase; LDH, lactic dehydrogenase; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); IPTG, isopropyl D-thiogalactopyranoside.

■ REFERENCES

(1) Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep. 18*, 380−416.

(2) Walsh, C. T. (2004) Polyketide and nonribosomal peptide antibiotics: Modularity and versatility. *Science 303*, 1805−1810.

(3) Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol. 7*, 285−295.

(4) Horinouchi, S. (2008) Combinatorial Biosynthesis of Non-bacterial and Unnatural Flavonoids, Stilbenoids and Curcuminoids by Microorganisms. *J. Antibiot. 61*, 709−728.

(5) Wong, F. T., and Khosla, C. (2012) Combinatorial biosynthesis of polyketides: A perspective. *Curr. Opin. Chem. Biol. 16*, 1−7.

(6) Fisch, K. M., Bakeer, W., Yakasai, A. A., Song, Z., Pedrick, J., Wasil, Z., Bailey, A. M., Lazarus, C. M., Simpson, T. J., and Cox, R. J. (2011) Rational Domain Swaps Decipher Programming in Fungal Highly Reducing Polyketide Synthases and Resurrect an Extinct Metabolite. *J. Am. Chem. Soc. 133*, 16635−16641.

(7) Ridley, C. P., Lee, H. Y., and Khosla, C. (2008) Evolution of polyketide synthases in bacteria. *Proc. Natl. Acad. Sci. U.S.A. 105*, 4595−4600.

(8) Zhang, F., He, H.-Y., Tang, M.-C., Tang, Y.-M., Zhou, Q., and Tang, G.-L. (2011) Cloning and Elucidation of the FR901464 Gene Cluster Revealing a Complex Acyltransferase-less Polyketide Synthase Using Glycerate as Starter Units. *J. Am. Chem. Soc. 133*, 2452−2462.

(9) Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., Chang, K. H., Liang, P. H., and Dunaway-Mariano, D. (1992) Ancestry of the 4-chlorobenzoate dehalogenase: Analysis of amino acid sequence identities among families of acyl:adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases. *Biochemistry 31*, 5594−5604.

(10) Gulick, A. M., Lu, X., and Dunaway-Mariano, D. (2004) Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states. *Biochemistry 43*, 8670−8679.

(11) Horswill, A. R., and Escalante-Semerena, J. C. (2002) Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: Residue Lys592 is required for propionyl-AMP synthesis. *Biochemistry 41*, 2379−2387.

(12) Kaneko, M., Ohnishi, Y., and Horinouchi, S. (2003) Cinnamate:coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3(2). *J. Bacteriol. 185*, 20−27.

(13) Lindermayr, C., Fliegmann, J., and Ebel, J. (2003) Deletion of a single amino acid residue from different 4-coumarate:CoA ligases from soybean results in the generation of new substrate specificities. *J. Biol. Chem. 278*, 2781−2786.

(14) Pohl, N. L., Hans, M., Lee, H. Y., Kim, Y. S., Cane, D. E., and Khosla, C. (2001) Remarkably Broad Substrate Tolerance of Malonyl-CoA Synthetase, an Enzyme Capable of Intracellular Synthesis of Polyketide Precursors. *J. Am. Chem. Soc. 123*, 5822−5823.

(15) Schneider, K., Hovel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., and Stuible, H. P. (2003) The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. *Proc. Natl. Acad. Sci. U.S.A. 100*, 8601−8606.

(16) Wu, R., Reger, A. S., Cao, J., Gulick, A. M., and Dunaway-Mariano, D. (2007) Rational redesign of the 4-chlorobenzoate binding site of 4-chlorobenzoate:coenzyme A ligase for expanded substrate range. *Biochemistry 46*, 14487−14499.

(17) Koetsier, M. J., Jekel, P. A., van den Berg, M. A., Bovenberg, R. A., and Janssen, D. B. (2009) Characterization of a phenylacetate-CoA ligase from *Penicillium chrysogenum*. *Biochem. J. 417*, 467−476.

(18) Hughes, A. J., and Keatinge-Clay, A. (2011) Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Func-tional Showcasing of Streptomyces coelicolor MatB. *Chem. Biol. 18*, 165−176.

(19) Beuerle, T., and Pichersky, E. (2002) Enzymatic Synthesis and Purification of Aromatic Coenzyme A Esters. *Anal. Biochem. 302*, 305−312.

(20) Yew, W. S., and Gerlt, J. A. (2002) Utilization of L-ascorbate by *Escherichia coli* K-12: Assignments of functions to products of the yjf-sga and yia-sgb operons. *J. Bacteriol. 184*, 302−306.

(21) Kim, Y. S., and Bang, S. K. (1988) Assays for malonyl-coenzyme A synthase. *Anal. Biochem. 170*, 45−49.

(22) Reger, A. S., Wu, R., Dunaway-Mariano, D., and Gulick, A. M. (2008) Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry 47*, 8016−8025.

(23) Kim, M. I., Kwon, S. J., and Dordick, J. S. (2009) In vitro precursor-directed synthesis of polyketide analogues with coenzyme A regeneration for the development of antiangiogenic agents. *Org. Lett. 11*, 3806−3809.

(24) McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., and Khosla, C. (1995) Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature 375*, 549−554.
(25) Horinouchi, S. (2009) Combinatorial biosynthesis of plant medicinal polyketides by microorganisms. *Curr. Opin. Chem. Biol. 13*, 197−204.
(26) Weeks, A. M., and Chang, M. C. Y. (2011) Constructing de Novo Biosynthetic Pathways for Chemical Synthesis inside Living Cells. *Biochemistry 50*, 5404−5418.
(27) Gerlt, J. A., and Babbitt, P. C. (2009) Enzyme (re)design: Lessons from natural evolution and computation. *Curr. Opin. Chem. Biol. 13*, 10−18.
(28) Tawfik, D. S. (2006) Biochemistry. Loop grafting and the origins of enzyme species. *Science 311*, 475−476.
(29) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605−1612.