

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Production of multi-methyl-branched fatty methyl ketones via a chimeric polyketide synthase pathway

Permalink

<https://escholarship.org/uc/item/2kj9j6m1>

Author

Curran, Sam C

Publication Date

2019

Peer reviewed|Thesis/dissertation

Production of multi-methyl-branched fatty methyl ketones via a chimeric polyketide
synthase pathway

by

Samuel Curran

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Comparative Biochemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D Keasling, Chair

Professor Michelle Chang

Professor Fenyong Liu

Fall 2019

Production of multi-methyl-branched fatty methyl ketones via a chimeric polyketide synthase pathway

Copyright © 2019
By Samuel Conrad Curran

Abstract

Production of multi-methyl-branched fatty methyl ketones via a chimeric polyketide synthase pathway

by

Samuel Conrad Curran

Doctor of Philosophy in Comparative Biochemistry

University of California, Berkeley

Professor Jay D Keasling, Chair

Petroleum and its derivatives revolutionized every aspect of human life, from how we work and travel to what we wear and eat. However, mitigation and prevention of climate change are of utmost ethical and economical importance. Biology is a potential means for reducing greenhouse gas emissions, while maintaining or even diversifying and improving the goods and services the world has grown to depend on. Biofuels, for example, can be made with renewable waste feedstocks and can supplement or replace petroleum-derived fuels but lack optimal molecular properties. We therefore sought a means for customizing biologically-produced carbon backbones. To this end, we engineered enzymes from microbial antibiotic biosynthesis called polyketide synthases. These enzymes form modular “assembly lines” where the activity of each catalytic unit can be predicted with reasonable confidence. We envisioned combining individual enzymes to form a new biosynthetic assembly line which, until this work, did not exist in nature. We first investigated the main biosynthetic chassis, an iterative polyketide synthase, BorM5, from the borrelidin pathway. We discovered that BorM5 is promiscuous for both starting and elongation substrates, producing a variety of mid-length branched and linear saturated fatty acyl intermediates. We next characterized the *trans*-acting thioesterase, BorB, hoping to use it as an offloading mechanism. We propose that BorB acts to remove dead-end intermediates from the borrelidin pathway, increasing its efficiency. We then engineered on- and off-loading neighbors for BorM5 and observed production of new-to-Nature C11 and C14 methyl branched methyl ketones. Diversification and high-titer production of these compounds, combined with materials testing, may lead to improved biofuels and the environmental, economic, and social benefits that accompany them.

Table of Contents

Table of Contents	i
Acknowledgements	iv
Chapter 1. Introduction	1
1.1. The petroleum revolution and the petroleum problem	1
1.2. Biofuels as a solution	2
1.3. Microbial production of advanced biofuels	4
1.3.1. Branched chain fatty acid biosynthesis	5
1.3.2. Isoprenoid synthesis.....	7
1.3.3. Polyketide synthases	8
1.4. A new-to-nature polyketide pathway for branched chain methyl ketones	9
Chapter 2. Polyketide synthase biology, diversity, and engineering	11
2.1. Biological roles of PKSs.....	11
2.2. Diversity of PKSs	12
2.2.1. Bacterial Type II iterative PKSs	13
2.2.2. Fungal Type I iterative PKSs.....	13
2.2.3. Type I modular Polyketide synthases (<i>cis</i> - and <i>trans</i> -AT).....	15
2.3. Biochemistry and engineering of Type I modular PKSs.....	17
2.3.1. Chain length	17
2.3.2. α -Keto substitution.....	20
2.3.3. β -keto reduction.....	21
2.3.4. Product release.....	23
2.3.5. Persistent issues and the cutting edge of PKS engineering research	23
Chapter 3. Probing the flexibility of an iterative modular polyketide synthase with non-native substrates in vitro	25
3.1. Abstract.....	25
3.2. Introduction	25
3.2.1. Demonstrating BorM5 activity in vitro	28
3.2.2. Expanding the library of non-native substrates.....	31
3.2.3. Engineering BorM5 to accept malonyl-CoA.....	34
3.3. Conclusion	35
3.4. Materials and methods.....	38
3.4.1. Reagents	38
3.4.2. Cloning.....	38
3.4.3. Acyl-CoAs.....	39
3.4.4. Acyl-SNAC/TP synthesis.....	39

3.4.5. NMR spectra.....	48
3.4.6. Protein purification	48
3.4.7. SDS-PAGE of purified proteins.....	50
3.4.8. LCMS sample run and data acquisition.....	50
3.4.9. LC-MS/MS Phosphopantetheine ejection assay	52
3.4.10. Raw data	53
3.4.11. Data analysis	53
3.4.12. Phosphopantetheinylation assays (acyl-ACP formation).....	54
3.4.13. Extension assays	54
3.5. Supporting Figures	55
3.6. Acknowledgements	63
3.7. Funding Sources.....	63
Chapter 4. Structure, function, and engineering of BorB, the type II thioesterase from the borrelidin biosynthetic cluster	64
4.1. Abstract.....	64
4.2. Introduction.....	64
4.3. Results.....	65
4.3.1. BorB is a TEII.....	65
4.3.2. BorB substrate specificity	66
4.3.3. Structure of BorB - overall structure	69
4.3.4. Structure of BorB - catalytic site	71
4.3.5. Structure of BorB - substrate binding chamber.....	71
4.3.6. Engineering BorB.....	73
4.4. Discussion	73
4.5. Methods.....	74
4.5.1. BorB accession numbers	74
4.5.2. TEII sequence scraping, filtering and alignment.	74
4.5.3. Plasmid Construction.....	77
4.5.4. Protein purification for biochemical assays.....	78
4.5.5. Protein expression for crystallography	80
4.5.6. Size-Exclusion Chromatography-Small Angle X-ray Scattering (SEC-SAXS).....	81
4.5.7. Crystallization, X-ray data collection and structure determination.....	84
4.5.8. Manual substrate docking.....	86
4.5.9. NMR	86
4.5.10. Acyl-SNAC synthesis.....	86
4.5.11. Turnover assays	92
4.5.12. Kinetic Assays	92
4.5.13. Kinetics derivatization	92

4.5.14. In vitro ACP-acylation and MALDI-TOF	93
4.6. Supporting Figures	94
4.7. Acknowledgements	106
4.8. Funding	106
Chapter 5. Engineering a chimeric borrelidin assembly line for methyl ketone production	107
5.1. Abstract.....	107
5.2. Introduction	107
5.3. Results.....	108
5.3.1. Engineering a loading module	109
5.3.2. Engineering an offloading module	112
5.3.3. Production and detection of methyl ketones in vitro.....	114
5.4. Discussion	118
5.5. Methods.....	119
5.5.1. Reagents	119
5.5.2. Acyl-SNAC synthesis.....	119
5.5.3. Acyl-CoAs.....	119
5.5.4. Enzyme domain boundaries	119
5.5.5. Plasmid Construction.....	119
5.5.6. Protein purification	120
5.5.7. In vitro assays.....	122
5.5.8. Protein mass spectrometry	123
5.5.9. GC-MS	123
5.6. Supporting Figures	124
5.7. Acknowledgements	126
5.8. Funding	126
Chapter 6. Conclusion	127
Chapter 7. References	130

Acknowledgements

First and foremost, I thank Laure Leynaud-Kieffer for her unfaltering support, patience, and love. I thank my parents, Thomas and Marcia Curran, for giving me the freedom to find my own path. I thank my friends for condolences and celebration as they were needed.

While this work is under my name, it could not have been accomplished without the help of my students, Andrew Lau, Marian-Joy Baluyot, Julie Lake, and Hendrik Pütz. They made significant and irreplaceable contributions to this work and to my maturation as a scientist and leader.

Equally, my collaborators and mentors made this work possible. Jay Keasling, Leonard Katz, Fenyong Liu, Satoshi Yuzawa, and Andrew Hagen provided the intellectual leadership for this thesis. Significant contributions were made by Sean Poust, Leanne Jade G. Chan, Brett Garabedian, Tristan de Rond, Christopher Petzold, Yan Chen, Jose Henrique Pereira, Daniel Rosenburg, Michal Hammel, Jim Kirby, Andria Rodriguez, and Jacquelyn Blake-Hedges.

Finally, I thank all the researchers before me whose diligent study informed my own work.

Chapter 1. Introduction

1.1. The petroleum revolution and the petroleum problem

Humans have exploited fossil fuels for millennia. Records of coal mining date to 3490 BC China and its use in metalworking is recorded as early as 250 BC Rome. (1, 2) Coal was the primary energy source for steam engines during the Industrial Revolution. Oil was first intentionally mined in Pennsylvania in 1859 and kerosene was first distilled from oil in that same year. (3) Liquid fuels, easily transported and dense in energy, permitted rapid and economical travel and commerce. Gasoline use coincided with the invention of the automobile in 1892. (4) 30 years later, nine million gasoline powered cars were on the road. Starting in the late 19th century, polymer chemists began discovering new materials. Petroleum-derived polymers such as polyethylene, polyethylene terephthalate (PET), conjugated dienes, and polyamides now dominate global markets as sources of plastics, rubbers, and fabrics. (5) This petroleum revolution initiated a new era of productivity and quality of life that continues to dominate our markets and daily lives.

This burgeon of productivity is not without its costs, however. The United Nations Intergovernmental Panel's Climate Change 2014 Synthesis Report summarizes records of unprecedented climate change since 1950. (6) The last three decades were likely the warmest in the last 800 years, with surface, upper ocean, lower atmosphere and polar ice core temperatures rising since 1970. The rise is attributed to the increase in greenhouse gasses (GHG) like carbon dioxide (CO₂) and methane (CH₄), which are the highest than they have been for at least the last 800,000 years. Human generated GHGs are calculated to be the single greatest contributor to radiative forcing, or the trapping of solar energy in the atmosphere, as well as the single largest contributor to the observed rise in global temperature. Between 1970 and 2012, 78% of these GHGs were produced through fossil fuel consumption and industrial processes.

Abiotic, biotic, and human systems are sensitive to climate change. Glaciers are consistently reducing in size, threatening water supply for communities that rely on meltwater, such as the San Francisco and Los Angeles metropolitan areas. (7) There has been an increase in droughts, fires, windstorms and pest outbreaks that are associated with climate change. While Nature can evolve over long time periods, the relative rapidity of anthropogenic climate change over the last century threatens ecosystems. Climate change has altered the distributions and activities of marine and terrestrial animals and has led to effects such as coral bleaching. Directly affecting humans, a net negative effect on agriculture has been reported, with reductions in crop yield and subsequent price increases. Extreme weather events have increased in frequency over the past decades with concomitant increases in direct and insured losses due to weather-

related disasters. These losses are felt most strongly by undeveloped countries who lack the infrastructure to mitigate and cope. However, developed countries feel the economic burden stemming from a reduction in ecosystem services, such as timber or fisheries production. A global increase of 4°C over the pre-industrial average is estimated to cost between 1-5% of global GDP per year. (8)

Thus, strategies that mitigate climate change are of economical and general value. Political and governmental recognition of this fact lead to the 2015 signing of the Paris Agreement pledge to decrease greenhouse emissions by 187 countries (96% of countries). (9) Apart from reducing consumption, we may alter the methods we employ in production; increasing efficiency, decreasing waste, or using alternative technologies. Fossil fuels, as a source for global commodities and transport, underpin global quality and means of life. As such, they are a prime target for innovations or replacements will lessen the emission of GHGs and mitigate global climate change and associated costs.

1.2. Biofuels as a solution

Prior to the petroleum revolution, biology was the main provider of materials and fuels. Whale oil (fatty acids and esters thereof) was used as an illuminant before the widespread adoption of kerosene. (10) Biology provided the materials for early polymer research; rubber was first produced with the terpene-rich “gum” of the equatorial tree *Hevea brasiliensis*. (5) Nylons, now produced with petrochemicals, have their roots in monomers derived from the castor bean (*Ricinus communis*). Styrene, the precursor for polystyrene foam, was first produced by distillation of storax balsam from the Mediterranean sweet gum tree. (5)

Biology has resurged for its capacity to provide fuels and materials. Biofuels are a popularly recognized group of biologically-produced molecules that are combusted for heat or energy. Two prominent examples are soy biodiesel and corn ethanol. Soy biodiesel accounted for over half of US biodiesel production in July 2019. (11) Fermentation of corn by yeast yields ethanol, which is added to gasoline in the United States at about 10% total volume. (12) While petroleum combustion leads to net GHG release, biofuels have the potential to generate a “partially closed-loop” cycle, where the GHG generated is offset by the initial GHG investment required to make the feedstock. GHGs emitted from the production and combustion of corn ethanol and soy biodiesel are 12% and 41% less than if fossil fuels were used to release the same amount of energy. (13) Importantly, ethanol and biodiesel yield 125% and 193% of the energy invested into their production (excluding solar).

However, the widespread adoption of fatty acid biodiesels is complicated by several issues. First, the fermentation of corn, soybeans, and other feedstocks for energy directly competes with food production. The use of other feedstocks, such as cellulosic

agricultural waste or plant matter grown on marginal land, may avoid competition and provide the GHG reductions necessary to combat climate change. (14) Cellulosic biomasses is comprised of non-food plant material, such as corn stalks and leaves, and is chemically composed of the polymers cellulose, hemicellulose and lignin. However, unlike soybean oil, these fibrous materials are cannot be easily converted to a suitable fuel.

The second group of issues surrounding biodiesels is that their chemical composition differs from the petroleum-derived diesel that our engines are designed for. Petroleum diesel is a complex mixture of up to thousands of chemical species, including linear (paraffins and olefins; alkanes and alkenes), cyclic (naphthenes), and aromatic hydrocarbons (Figure 1-1). (15) The average number of carbons in commercial D-2 diesel fuel is 16. Cetane number (CN) is the primary quality score of a diesel fuel and is inversely related to ignition delay. Higher CNs are therefore generally indicative of better fuels. The straight-chain hydrocarbon heptadecane has a CN of 100. (16)

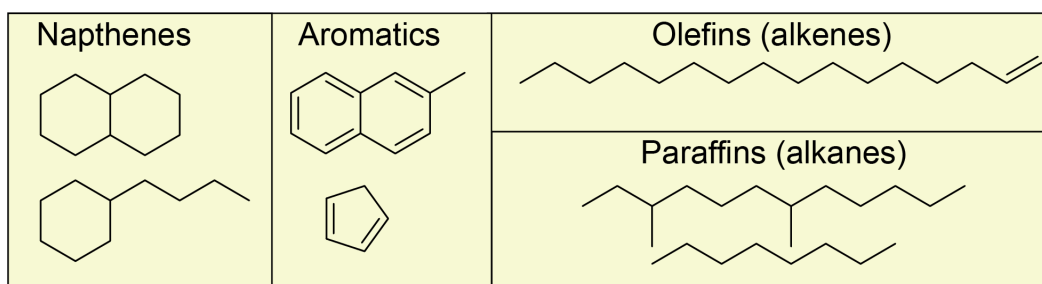


Figure 1-1: Examples of biodiesel components.

Biodiesels are produced from plant-derived fatty acids. (17) The length of the hydrocarbon and the degree of saturation vary significantly with the crop used; soy-based biodiesel is predominantly C18:1 while mustard-based biodiesel is distributed over C16-C22 with varying degrees of saturation. (18) The carboxylic acid moiety of biodiesels leads to several issues such as a high melting point (63°C for hexadecanoic acid), high viscosity, and high hygroscopicity, or sequestration of water. (19) For these reasons, the fatty acids are esterified to methyl- and ethyl esters (Figure 1-2). Esters of hexadecanoic acid have CN numbers of 80 or higher. (19)

Nevertheless, the low freezing point and biodiesels remains an issue; the methyl ester of hexadecanoic acid freezes at 30.5°C which is unsuitable for temperate climates. (16) Several strategies exist for depressing freezing points, but these have caveats. Decreasing chain length can decrease freezing point, but also decreases CN. Desaturating fatty acids decreases the freezing point but increases viscosity and decreases oxidative stability (Figure 1-2). One underexplored method for decreasing freezing point is chain branching. Terminally-branched (iso) fatty acids have lower freezing points than their

linear counterparts. (20) The position of the branch also affects the melting point; a methyl-branch two carbons from the terminus (anteiso) yields a greater freezing point reduction than an iso-branched fatty acid. One downside is that chain branching can reduce CN. Heptamethylnonane, for example, has a CN of 15. (16) However, this chemical space is largely unexplored due to the biological scarcity of branched-chain fatty acids. Suitable biodiesel molecules may yet exist, although the current feedstocks are incapable of producing them.

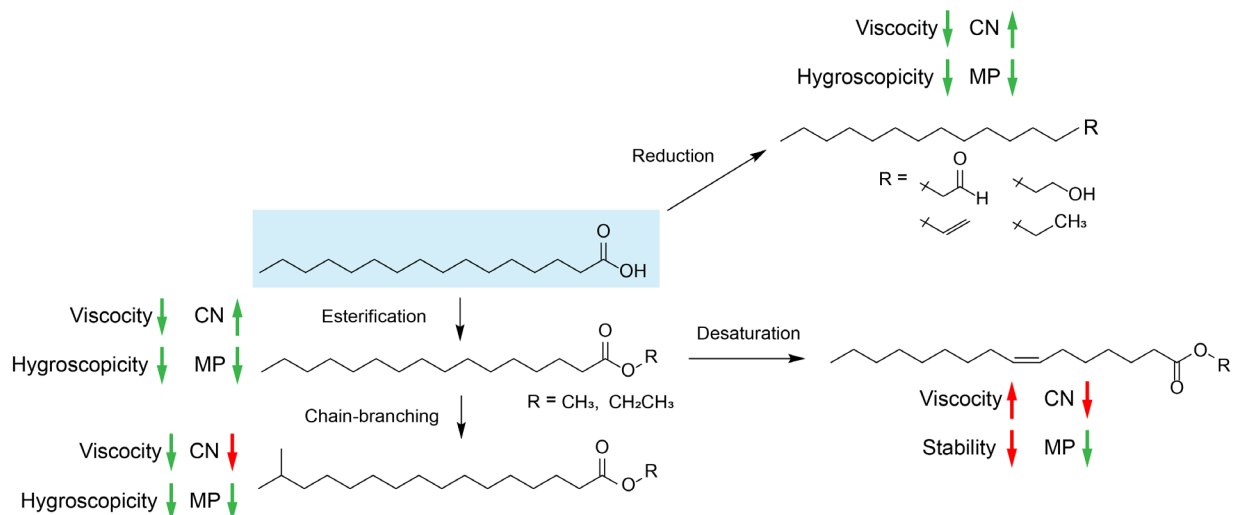


Figure 1-2: Relationship between molecular structure and fuel properties.

1.3. Microbial production of advanced biofuels

Microbial engineering provides a means to convert cellulosic biomass to energy-dense advanced biofuels, such as branched-chain fatty acids and their derivatives. Microorganisms such as bacteria fungi natively degrade plant polymers and metabolize the subunits for growth, energy storage and reproduction. Biological anabolic reactions, or the “building up” of molecules, offer exquisite control of stereo- and regiochemistry, the use of simple starting materials, and do not require hazardous reactive reagents. Our ability to engineer biology was limited until the early 2000s when the human genome was sequenced. (21) The ability to derive and analyze sequence data, among other innovations, heralded the field of “synthetic biology”, a combination of classical biology and engineering disciplines. (22) Synthetic biology practitioners leverage computational and engineering tools to understand and engineer bacterial metabolism. Cutting edge efforts include the understanding of diverse organisms and their biosynthetic pathways, the engineering of control systems, and the development of “new-to-Nature” biosynthetic pathways to reach non-native chemical space.

Biological pathways have been engineered and leveraged for economically-relevant fatty acids, alcohols, amines, alkanes, amino acids, and aromatics. (23) Additionally, several methods exist for the microbial production of saturated branched-chain fatty acids (BCFAs) and their derivatives. These will be discussed below.

1.3.1. Branched chain fatty acid biosynthesis

Fatty acid synthesis (FAS) is a ubiquitous anabolic pathway, found in archaea, bacteria, plants and animals (although variations exist among the structures, substrates, uses and regulation). Fatty acids form cellular membranes in prokaryotes and eukaryotes. Archaea have a fatty alcohol-glycerol membrane but nevertheless use fatty acids to acylate proteins. (24) FAS consists of the sequential carbon-carbon forming Claisen-like condensation of enzyme-bound malonate units and the growing intermediate. Intermediates are covalently tethered to the active sites of the enzyme, typically via thioester linkages. After elongation, the β -keto intermediate is fully reduced and the cycle continues until a certain chain length is reached.

FAS in diverse hosts shares a sequence of reactions and domain usage. However, the architectures of the various enzymes that perform these reactions are wildly different. Mammalian FAS is housed on a 270kDa polypeptide that contains all the domains necessary to synthesize fatty acids and forms dimers. (25) Fungal FAS is broken up into condensing and reducing \square and \square chains, which form a 2.6 megadalton spherical dodecamer. (26) On the other hand, bacteria such as *Escherichia coli* have FAS systems where the catalytic domains are on separate polypeptide chains. (27) These 15-50 kDa domains form non-covalent assemblies to synthesize fatty acids.

The generalized catalytic cycle of FAS is depicted in Figure 1-3. Prior to biosynthesis, the acyl-carrier protein (ACP) is post-translationally modified via the covalent linkage of phosphopantetheine moiety at the active site serine. The “phosphopantetheine arm” adds 20 Å to the reach of the ACP, permitting it to enter the active sites of the neighboring domains. Biosynthesis is primed with acetyl-CoA, and malonyl-CoA derived from the TCA cycle (Figure 1-3). (27) Malonate is loaded onto an ACP via a malonyl acyltransferase (MAT) and then transacylates onto the ketosynthase (KS) domain active site cysteine. In *E. coli*, malonyl-CoA is biosynthesized via the carboxylation of acetyl-CoA by the biotin dependent acetyl-CoA carboxylase complex. (27) The KS catalyzes a decarboxylative Claisen-like condensation reaction with the extender unit, releasing and extending the priming substrate by two carbons and forming a β -keto acyl-ACP intermediate. *E. coli* utilizes three KSs (FabB, FabF, and FabH) each with a distinct chain-length specificity. (27) The extended β -keto intermediate is reduced sequentially by the NADPH-dependent ketoreductase (KR), dehydratase (DH), and NADPH-dependent enoyl reductase (ER).

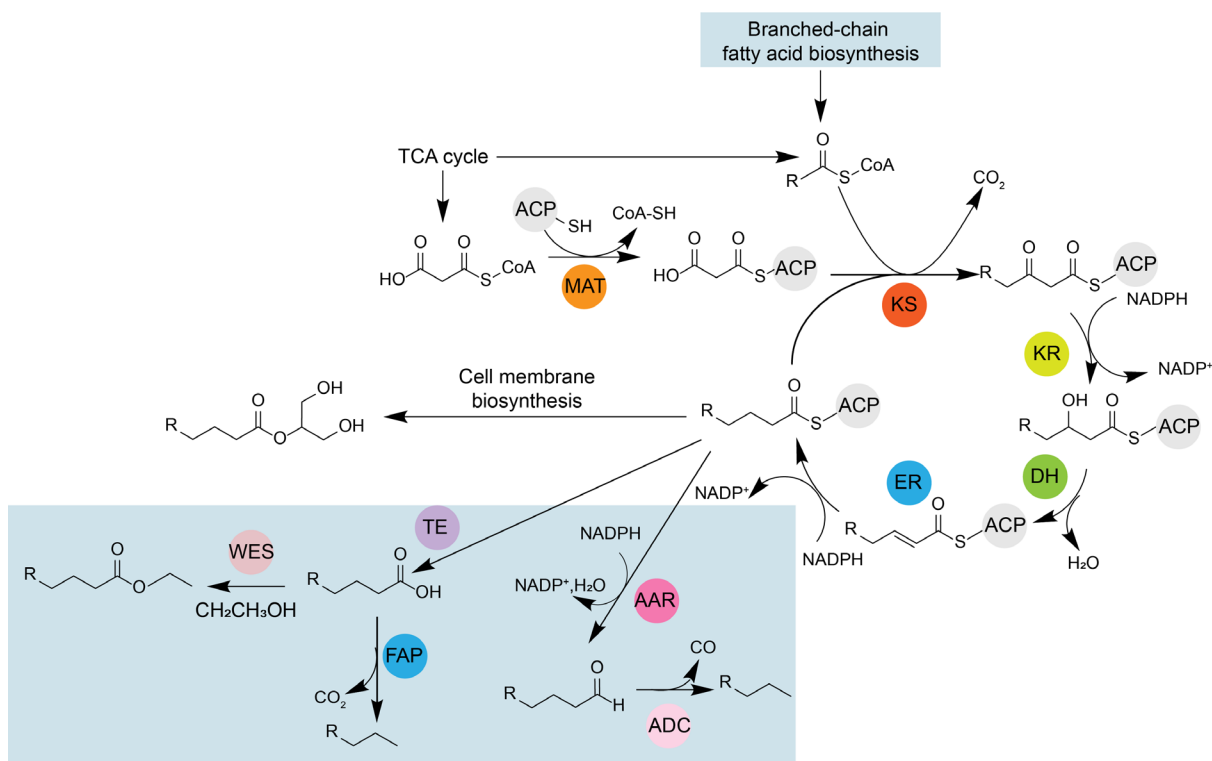


Figure 1-3: Native and engineered fatty acid biosynthesis in *E. coli*. Blue boxes delineate engineered biosynthesis in *E. coli*. ACP, acyl-carrier protein; MAT, malonyl acyltransferase; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoylreductase; TE, thioesterase; WES, wax ester synthase; FAP, fatty-acid photodecarboxylase; AAR, acyl-ACP reductase; ADC, aldehyde decarbonylase.

After reduction, the chain is either (1) transacylated to the KS domain for a subsequent round of elongation and reduction, or (2) released if it is sufficiently long to be recognized by chain-terminating enzymes. In *E. coli*, fatty acids are released as an acyl-phosphate or an acyl-glycerate. (27) Both of these compounds can be used in membrane biosynthesis. Free fatty acids can be produced by the expression of periplasmic enzyme thioesterase I (TesA) lacking a signal peptide ('TesA). This enzyme hydrolyzes fatty acyl-ACP in the cytoplasm, releasing ACP and free acids. (28)

Many Gram positive bacteria natively produce BCFAs (29). BCFA biosynthesis begins with the branched-chain amino acids leucine, isoleucine and valine (Figure 1-4). Aminotransferases convert these to α -keto acids, which are then used decarboxylated and converted to acyl-CoAs via the branched chain α -keto acid complex (BCKD). A dedicated ketosynthase primes fatty acids biosynthesis with these substrates and subsequently produces medium-chain iso- and anteiso- fatty acids via repetitive malonyl-CoA extension and reduction cycles.(30) *E. coli* lacks both the BKCD complex and a dedicated

KS and therefore does not natively produce BCFAs. However, expression of the *Bacillus subtilis* BCKD and KSs lead to the biosynthesis of iso- (~23 mg/L) and anteiso fatty acids (~150 mg/L) (Figure 1-4). (31, 32) Additional optimization of the BCKD pathway lead to an engineered strain where BCFA constituted 85% of total free fatty acids (276 mg/L). (33)

Fatty acids by themselves must be esterified before they exhibit desirable fuel properties. (16) Expression of a wax-ester synthase gene, *atfA*, in a fatty-acid overproducing strain, lead to the production of fatty acid ethyl esters when ethanol was exogenously provided (Figure 1-3). (34) The carboxylic acid can also be reduced; expression of the *Acinetobacter calcoaceticus* BD413 fatty acid reductase *acr1* produced fatty alcohols. (34) Reduction to aldehydes and alkanes is possible through the expression the acyl-ACP reductase and aldehyde decarbonylase (Figure 1-3). (35, 36) Alkane production is also attainable through direct fatty acid decarboxylation via a recently discovered algal photoenzyme. (37) Finally, methyl ketones can be generated through the decarboxylation of β -keto fatty acids. (38–43)

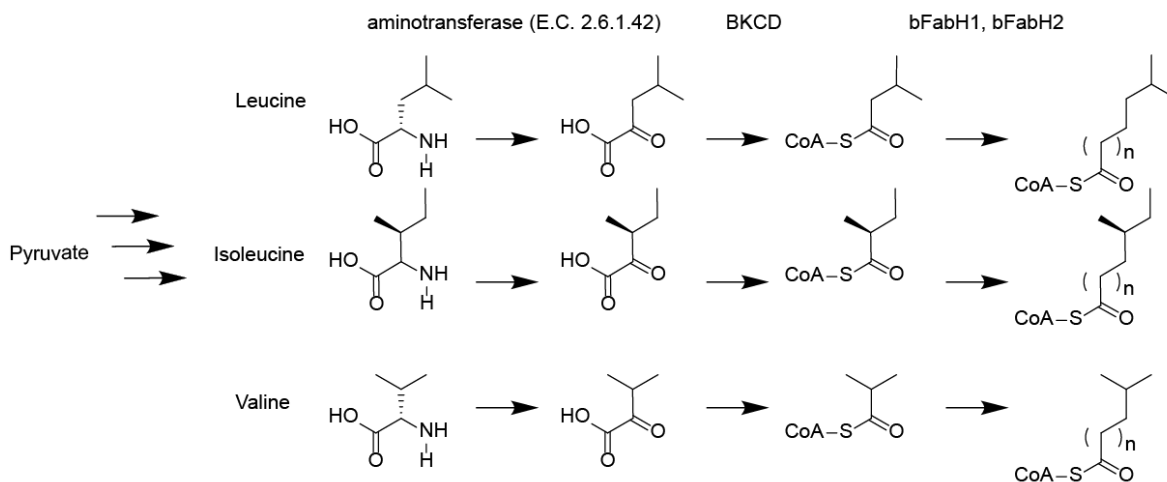


Figure 1-4: Branched chain fatty acid biosynthesis.

1.3.2. Isoprenoid synthesis

The isoprenoid family contains over 40,000 known structures and include pharmaceuticals, fragrances and flavors. (44, 45) Plants produce diverse isoprenoids but at insufficient titers for biofuels. For that reason, *E. coli* and the yeast *Saccharomyces cerevisiae* have been engineered to produce isoprenoids. (46, 47) Isoprenoids are derived from the 5-carbon precursors isopentenyl-diphosphate (IPP) and dimethylallyl-diphosphate (DMAPP), which are biosynthesized by either the mevalonate (eukaryotes

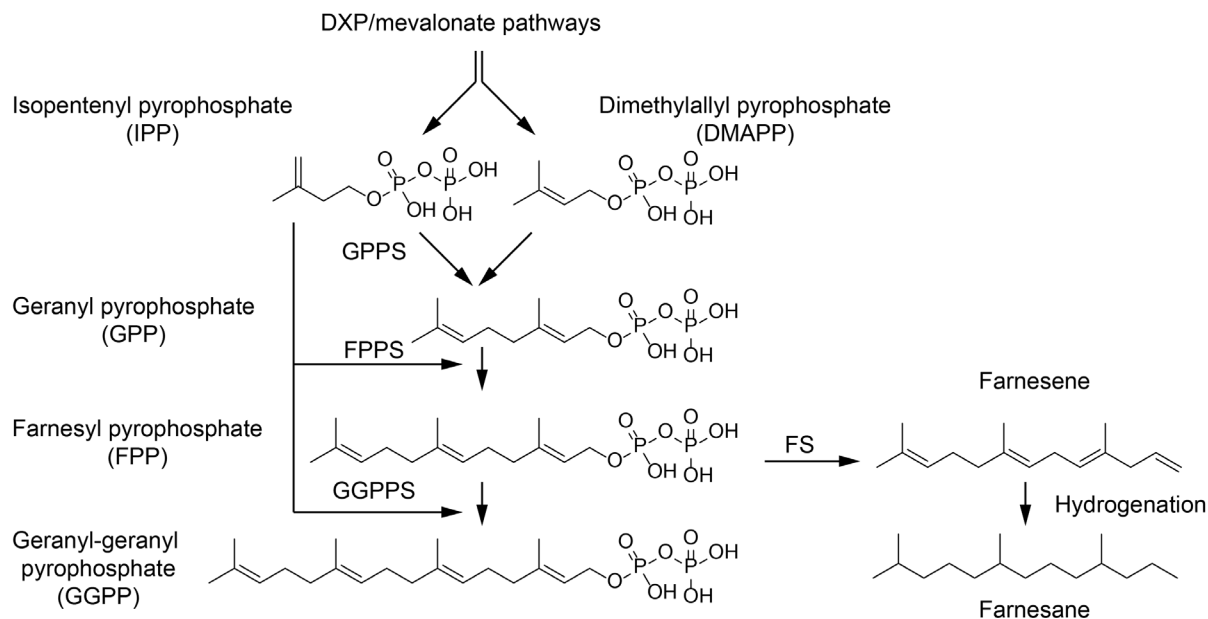


Figure 1-5: Farnesane (bio)synthesis. GPPS, geranyl pyrophosphate synthase; FPPS, farnesyl pyrophosphate synthase; GGPPS geranyl geranyl pyrophosphate synthase; FS, farnesene synthase

and some Gram positive bacteria) or deoxyxylulose-5-phosphate pathway (DXP; prokaryotes) (Figure 1-5). Prenyltransferases catalyze the head-to-tail addition of IPP and DMAPP units, leading to 10-, 15-, 20-, or 30-carbon mono-, sesqui-, di-, and triterpenes, respectively. Farnesane, the reduced form of the branched 15-carbon sesquiterpene farnesene, is currently marketed as a jet-fuel additive. (48)

1.3.3. Polyketide synthases

Polyketide synthases (PKSs) provide an additional means for producing branched chain fatty acids and their derivatives, as well as many diverse molecules. As this dissertation presents the engineering of a PKS pathway, these enzymes will be lightly introduced here, followed by Chapter 2 which provides a technical introduction to PKS biology and engineering.

Polyketide synthases and FAS share a common evolutionary ancestor and much of the same chemistry. (49) Indeed, the minimal PKS consists of the domains required for a chain elongation; the AT, ACP, and KS. However, a glance at the structures of polyketides not immediately lead to this conclusion. Polyketides are often larger than standard primary metabolites, contain numerous heteroatoms, chiral centers and rings, and can mediate biological interactions such as resource competition.

Polyketide biosynthesis is most easily understood through the lens of Type I modular PKSs. These enzymes are characterized by the use of multiple singly-extending “modules” that form biosynthetic “assembly lines”. Each module performs a single elongation and modification before passing the substrate downstream. The assembly line nature of these synthases simplifies the prediction of the product, which can now be at least partially predicted from sequence alone. (50)

Polyketide biosynthesis uses diverse starting and extension substrates compared to FAS. The use of branched-chain starting substrates and branched extension substrates, such as methylmalonyl-CoA, leads to the production of branched-chain fatty acids. PKSs have been used to produce branched long-chain esters and branched short-chain ketones (Figure 1-6). (42, 51) However, due to the difficulties of engineering these synthases, this chemical space remains largely unexplored.

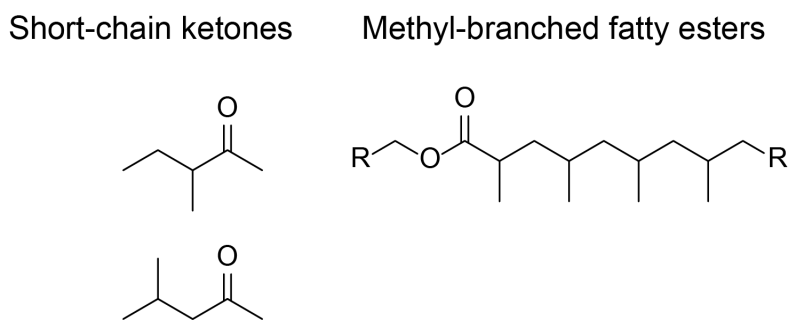


Figure 1-6: Short-chain ketones and methyl-branched esters produce via engineered PKS pathways.

1.4. A new-to-nature polyketide pathway for branched chain methyl ketones

To access a diverse array of methylated fatty acids (and their derivatives), we envisioned the construction of a new-to-Nature polyketide pathway to produce fatty acids and derivatives thereof with tunable methylation patterns. As an introduction to the field of PKS engineering, the biology, diversity, and engineering successes for these enzymes are reviewed in Chapter 2.

The forthcoming de novo pathway centers around a non-canonical mixed modular/iterative PKS, BorM5. (52) BorM5 natively extends intermediates with methylmalonyl-CoA and fully reduces the β -keto moiety. Natively, BorM5 is part of a heptaketide megasynthase that produces the aminoacyl tRNA synthetase inhibitor, borrelidin. BorM5 acts iteratively within its assembly line, extending the biosynthetic

intermediate three times. Chapter 3 details the in vitro characterization of BorM5 and its ability to synthesize diverse branched fatty acyl-ACPs.

To offload intermediates from BorM5 we investigated the hitherto uncharacterized *trans*-acting type II TE BorB to hydrolyze intermediates from BorM5. The structural and kinetic investigation of BorB is detailed in Chapter 4. BorB did not yield the desired products but its supporting role as an editing enzyme will be discussed. Chapter 5 details the in vitro and in vivo demonstrations of a full assembly line for

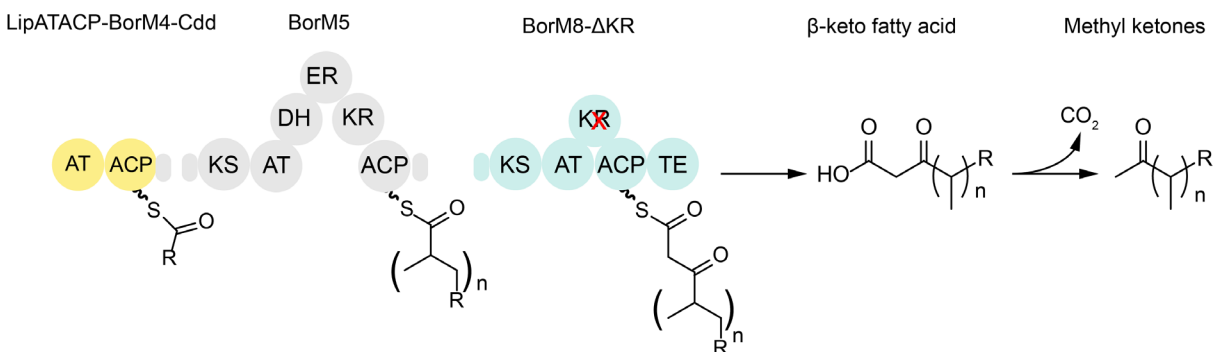


Figure 1-7: Engineered borrelidin pathway to produce methyl ketones.

methylated methyl ketone production (Figure 1-7). The assembly line uses the promiscuous lipomycin loading di-domain, which primes biosynthesis with short branched- and straight chain acyl-CoAs. (53) BorM5 is used as the core biosynthetic chassis, producing methylated fatty acyl-ACP. Intermediates are passed to a second module, BorM8, where they are offloaded as β-keto acids. These decarboxylate non-enzymatically to yield CO₂ and methyl-branched methyl ketones (MBMKs). (38–43)

Chapter 2. Polyketide synthase biology, diversity, and engineering

Contains material original co-authored as Barajas, J.F, Blake-Hedges, J.M., Bailey, C.B., Curran, S.C., and Keasling, J.D. Engineered polyketides: Synergy between protein and host level engineering. *Synthetic and Systems Biotechnology*. 2017. 2(3). 147-166. <https://doi.org/10.1016/j.synbio.2017.08.005>. Copyright Elsevier 2017. Reprinted with explicit co-author permissions and following publisher guidelines.

2.1. Biological roles of PKSs

Polyketides fit into a larger umbrella of compounds termed “Natural Products” (NPs) which also includes non-ribosomal peptide synthases (NRPSs), (54) ribosomally synthesized and post-translationally-modified peptides (RiPPs), (55) and isoprenoids, among others. PKSs are found in bacteria, fungi, plants, and even rarely in animals. (56) Soil-dwelling actinobacteria, such as the streptomycetes, are the canonical producers of polyketide Natural Products. In the “Golden Age of Antibiotics”, ranging from the 1950s to the 1970s, numerous antibiotics were isolated from cultures of streptomycetes. (57) The genome sequencing era, starting in the early 2000s, revealed the presence of dozens of unexpressed “cryptic” biosynthetic gene clusters (BGCs) in Natural Product hosts. For example, the sequencing of the *Streptomyces coelicolor* A3(2) genome revealed over 20 gene clusters for predicted Natural Products. (58) Modern metagenomic sequencing of uncultured environmental samples revealed massive diversity within the actinobacteria as well as other phyla, including acidobacteria and proteobacteria. (59) Established ecological roles for natural products include chemical defense, offense, and nutrient scavenging. (60) NPs produced or inspired roughly 50% of approved drugs between 1981 and 2014. (61) Pharmaceutically relevant polyketides include erythromycin A, lovastatin, and doxorubicin (Figure 2-1).

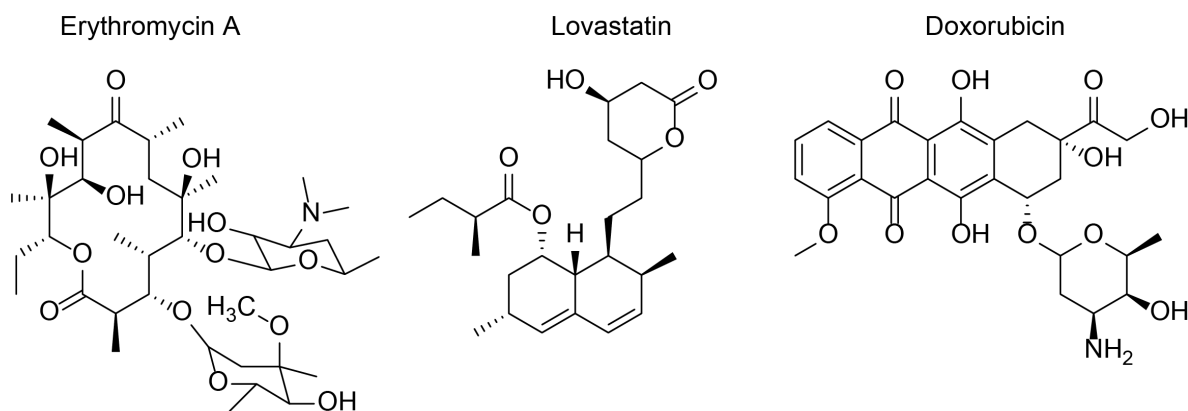


Figure 2-1: Structures of erythromycin A, lovastatin, and doxorubicin.

2.2. Diversity of PKSs

Phylogenetic analysis supports that PKS and FAS are derived from a common ancestor. A phylogenetic tree based on KS alignments shows the path that these enzyme classes have taken (Figure 2-2). (49, 62) The PKSs subclasses are differentiated by quaternary structure as well biosynthetic paradigm. Several relevant classes are discussed below.

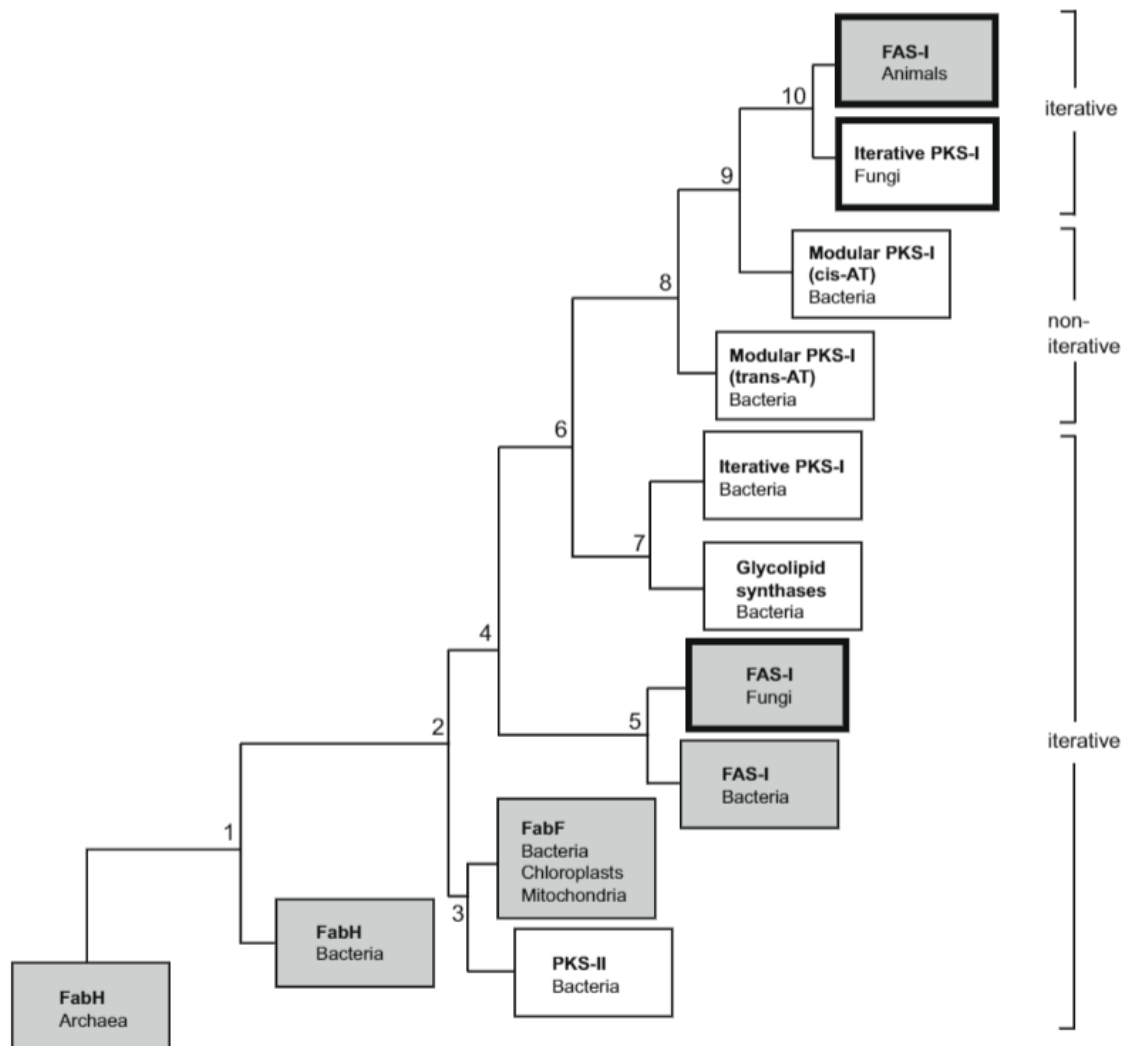


Figure 2-2. Phylogenetic tree of KS sequences from FASs and PKSs. This research was originally published as Jenke-Kodama H, Dittmann E (2009) Evolution of metabolic diversity: insights from microbial polyketide synthases. *Phytochemistry* 70(15-16):1858–1866. DOI: 10.1016/j.phytochem.2009.05.021. Copyright © Elsevier 2009. Reprinted with permission.

2.2.1. Bacterial Type II iterative PKSs

Like the *E. coli* FAS, Type II PKSs are iterative synthases that are comprised of dissociated enzymatic domains. These synthases typically produce multicyclic aromatic polyketides such as actinorhodin or tetracenomycin (Figure 2-3). (63, 64) The typical elongating domains are the ketosynthase- α (KS_{α}), KS_{β} or chain length factor (CLF), and acyl-carrier protein (ACP). (65) Priming occurs via either a dedicated priming module, such as in the synthases for the R1128 estrogen receptor antagonist or daunorubicin, (66) or by decarboxylation of malonyl-CoA to acetate by the CLF domain. (67) Chain elongation is catalyzed by the KS_{α} domain. The CLF lacks the catalytic cysteine and therefore cannot elongate intermediates. (68) However, the CLF plays the crucial role of limiting the chain length of the intermediate. Cyclization of the intermediates is mediated by cyclase (CYC) domains, which are thought to guide reactive poly- β -keto intermediates towards the correct cyclization pattern. As evidence for this, expression of the *whiE* minimal type II PKS (KS_{α} , KS_{β} , AT, ACP) without the CYC domain lead to the detection of a multitude of products. (69) Additionally, in actinorhodin biosynthesis, the shunt products SEK4b and SEK4 are detectable when the tailoring domains are not expressed. (70)

2.2.2. Fungal Type I iterative PKSs

A single fungus may have dozens of NP-BGCs; *Aspergillus niger* contains 34 different clusters. (71) Fungal iterative PKSs are most phylogenetically similar to mammalian FAS (Figure 2-2). (62) These systems are large multi-domain proteins wherein each domain has a distinct catalytic activity. (72) The domains collaborate to iteratively extend and

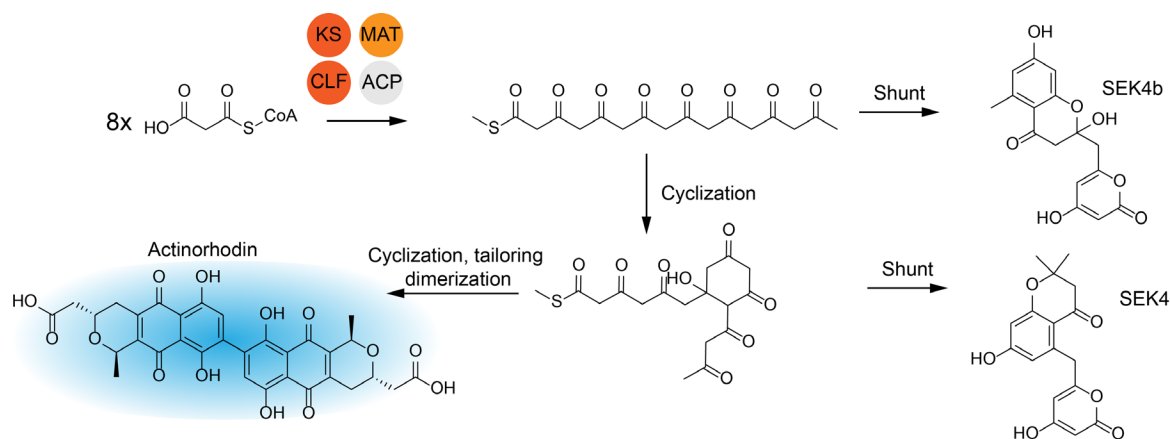


Figure 2-3: Blue pigment actinorhodin biosynthesis. This work was originally published as Keatinge-Clay AT, Maltby DA, Medzihradzky KF, Khosla C, Stroud RM (2004) An antibiotic factory caught in action. *Nat Struct Mol Biol* 11(9):888–893. DOI: 10.1038/nsmb808. Copyright © Springer Nature 2004. Adapted with permission.

modify acyl-intermediates. However, fungal iterative PKSs contain domains not found in FAS that mediate reactions such as cyclization, C-methylation, and amidation. (72–75)

These PKSs produce compounds that range from linear chains to complex multicyclic aromatics. (72–75) The first fungal PKS to be studied was the 6-methylsalicylic acid (6-MSA) synthase. (76) Since, the known members of this clade have grown to include the synthases for commercially important compounds such as the cholesterol-lowering lovastatin, the common agricultural contaminant aflatoxin, and the antifungal strobilurin. (72) Fungal iterative PKSs clade into three groups based on KS sequence and substrate selectivity: the non-reducing (NR), partially-reducing (PR), and highly-reducing (HR) fungal iterative PKSs. (77, 78). NR-PKSs are perhaps the simplest of these synthases. (72–75) PksA from aflatoxin biosynthesis is the best-studied example from this class (Figure 2-4). (75) The functions of individual domains were determined by individually expressing and combining them in vitro. (79) A starter unit S-acyl transferase (SAT) domain selects the starting unit (hexanoyl in the case of PksA) and provides it to the KS domain. (80) The malonyl-CoA:ACP transacylase (MAT) domain selects the extension substrate, which is condensed with the priming substrate by the KS domain. PksA iterates seven times to form an octaketide which is then cyclized. A structure of the

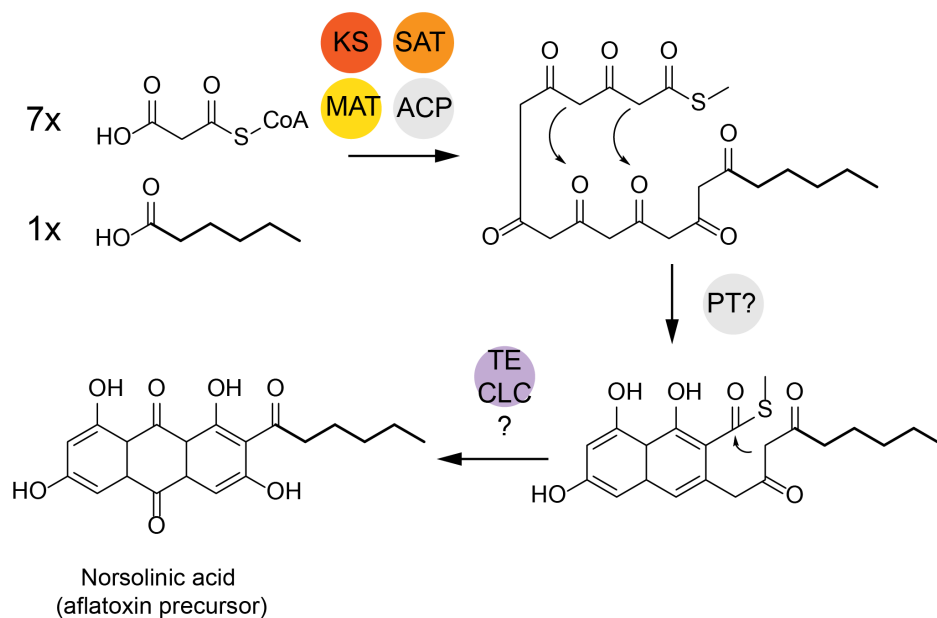


Figure 2-4: Aflatoxin polyketide core biosynthesis. The starting substrate, hexanoate, is in bold. This research was originally published as Crawford JM, Townsend CA. (2010). New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* 8(12):879–889. DOI: 10.1038/nrmicro2465. Copyright © Springer Nature 2010. Adapted with permission.

SAT-KS-MAT is available for the NR-PKS CTB1, which produces the polyketide backbone of cercosporin. (81, 82)

The mechanisms for selective activation and cyclization have been elucidated through structure-guided mutagenesis. (83, 84) The product template (PT) domain guides the cyclization of the reactive poly- β -keto intermediates. Swapping of PT domains can lead to novel cyclization patterns. (85) Additionally, NR-PKSs often contain a Claisen-like condensation thioesterase (CLC-TE) domain which is responsible for the final carbon-carbon cyclization. (75, 79) A crystal structure is available for the PksA CLC-TE domain. (86)

2.2.3. Type I modular Polyketide synthases (*cis*- and *trans*-AT)

The Type I PKSs are the focus of this dissertation. These synthases adopt biosynthetic assembly lines where the number of singly elongating modules determines the size of the product. The Type I modular synthases fall into two clades: *cis*-AT and *trans*-AT.

The *cis*-AT modular PKSs are the “textbook” PKSs. The best-studied example of the *cis*-AT PKSs is the 6-deoxyerythronolide B synthase (DEBS), which produces the aglycone core of the antibiotic erythromycin in the actinomycete *Saccharopolyspora erythraea* (Figure 2-5). (87) Sequencing of the DEBS gene cluster revealed three large open reading frames, *eryAI*, *eryAII* and *eryAIII*, coding for a loading di-domain and six polyketide modules split over three large polypeptides. (88) The coding regions were arrayed in the same order as their biosynthetic operations, a pattern described as “co-linearity”. The AT-ACP loading di-domain primes biosynthesis with propionate. The six modules load extend propionate with methylmalonate sequentially to produce a heptaketide. The reduction at each stage correlates with the presence or absence of functional reducing domains (except for the KR⁰ of DEBS-M3, which is redox inactive). For example, the full DH-ER-KR loop of DEBS-M4 reduces the β -ketone to a methylene. After all modules have extended, the intermediate is cyclized using the terminal hydroxyl moiety and released by the TE domain. Finally, the liberated macrocycle is hydroxylated and glycosylated by non-PKS “tailoring” enzymes. Due to its historical precedence, DEBS is one of the best studied polyketide synthases. (89) Much of our structural and biochemical understanding is derived from studies on these modules. These will be described in detail in section 2.3.

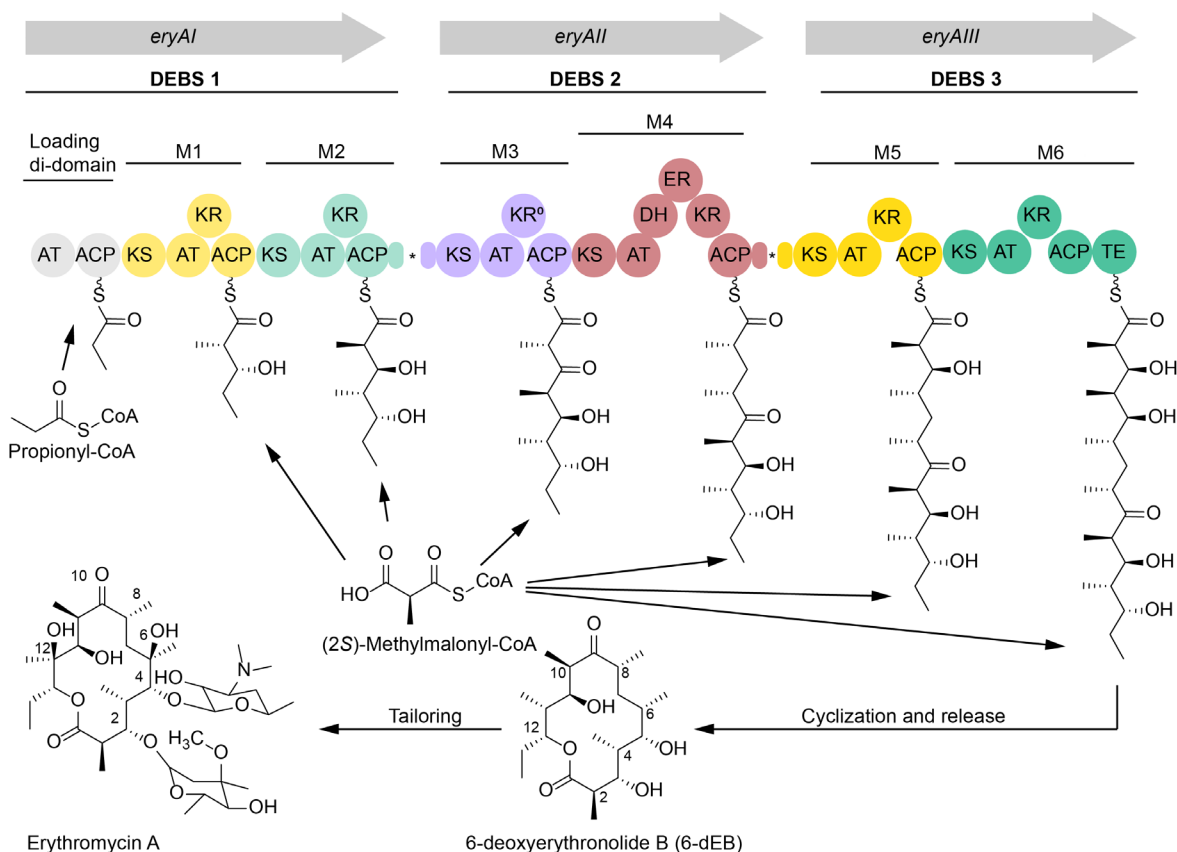


Figure 2-5. 6-deoxyerythronolide B synthase. This research was originally published in the Journal of Biological Chemistry. Cane D.E. Programming of Erythromycin Biosynthesis by a Modular Polyketide Synthase. J. Biol. Chem. 2010; 285:27517-27523. Copyright © the American Society for Biochemistry and Molecular Biology. Adapted with permission.

Trans-ATs PKSs, on the other hand, are less well understood. In these assembly lines, the AT domains are housed on separate polypeptides that interact in non-covalently with the main synthase (Figure 2-6). *Trans*-AT PKSs have diverse modular architectures, with duplicated or defunct domains, partial modules, non-elongating modules, and other domains that act in *trans*. (90) *Trans*-AT PKSs also use diverse domains outside of the standard elongation-reduction repertoire to catalyze reactions such as oxygen insertion in the pederin synthase, (91) Michael-branching in the rhizoxin synthase, (92) and pyran synthesis in bryostatin. (93) The *trans*-AT PKS are the less-well understood of the two types, despite accounting for upwards of 30% of identified PKSs. (90) Current efforts include using *trans*-AT components in engineered *cis*-AT assembly lines.

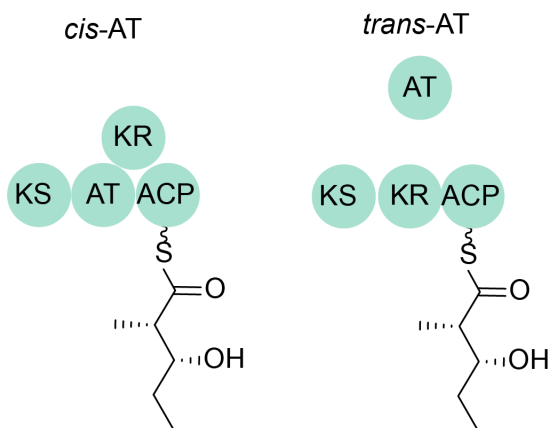


Figure 2-6. *cis*- and *trans*-AT PKS modules.

2.3. Biochemistry and engineering of Type I modular PKSs

The modular nature of these enzymes has inspired methods to mix and re-combine modules to produce new compounds. A recent study calculated that 20 of the 150 commodity chemicals that are followed by market tracking provider ISIS could be hypothetically produced by engineered PKSs. (94) PKSs have been used for the production of alkenes, (95) adipic acid, (96) methyl-branched esters, (97) hydroxy acids, (53) carboxylic acids and pyrones, (98) and short-chain ketones. (42) Strategies for engineering PKS assembly lines for custom molecule chain length and substitution include module recombination, domain swapping, and active site mutagenesis. (99) These will be highlighted below followed by a discussion of the challenges and paths forward in this developing field.

2.3.1. Chain length

Chain length is controlled by the number of elongations performed. In the Type I modular PKSs (in most cases), this is determined by the number of modules and begins with the loading module. PKSs can be primed with diverse compounds including short, medium and long chain saturated fatty acids, diacids, cyclic aliphatic acids, and more. (100) PKS loading modules have been split into two groups based on their architecture and biosynthetic sequence; about half of known loading modules contain non-decarboxylative KS^Q, AT, and ACP domains while the other half are comprised of an AT and an ACP (Figure 2-7). (67) The active cysteine of the KS^Q domain is replaced by glutamine (Q). Thus, these domains are incompetent for the condensation reaction but retain their decarboxylative activity. (67) The KS^Q decarboxylates malonyl- or

methylmalonyl-CoA yield acetyl- or propionyl units, respectively. (101) The AT domains of these modules are strictly specific for dicarboxylic acid substrates (malonyl- or methylmalonyl-CoA) (101) and share a conserved arginine residue with extender AT domains that is used to stabilize the free carboxyl moiety of the substrate. (102) Mutation of this residue to a tryptophan in the oleandomycin synthase afforded a mutant that could prime biosynthesis with propionate and acetate, albeit at 1% yield of the native module. (101)

Non-decarboxylative LDDs prime polyketide biosynthesis with a much broader range of substrates including linear and branched acids and diacids. Natively, the avermectin LDD primes biosynthesis with either 2-methylbutyryl-CoA or isobutyryl-CoA. However, feeding studies that the avermectin synthase could use over 40 different carboxylic acid starting substrates. (103) Similarly, while the related lipomycin LDD primes biosynthesis with isobutyryl-CoA *in vivo*, it was demonstrated to be promiscuous *in vitro*, loading a variety of branched and straight-chain reduced carboxylic acids. (53) Conversely, the borrelidin LDD is selective for dicarboxylic acids both *in vivo* (104) and *in vitro*. (105)

The structural basis for substrate selectivity, or lack thereof, by the non-decarboxylative LDDs was recently partially elucidated with an X-ray crystal structure of the avermectin LDD AT monodomain, AveAT0. (106) Notably, this domain contains a tryptophan in place of the conserved arginine of decarboxylative loading modules. Modeling of substrate into a putative binding tunnel identified targets for modulating substrate recognition. Sequence alignment of AT domains from KS^o loading domains and LDD demonstrated that these residues are less conserved in LDDs, correlating with their greater range of observed priming substrates. Individual mutagenesis of these residues resulted in a mutant (V222L) with ~30-fold decreased activity for the native substrate while retaining activity for pentanoyl-CoA. (106) Nevertheless, visualization of the AveAT0 binding pocket provides a strong starting point for future mutagenesis experiments combined with better measures of substrate specificity to increase catalytic activity toward individual substrates.

Chain elongation is mediated by KS domains. Historically, the role of the KS domain in moderating substrate specificity in type I *cis*-AT PKSs has been unclear. However, type I *trans*-AT PKSs show clear relationships between the primary KS amino-acid sequence and the substrate. In an alignment of *trans*-AT modules, KS domains clade not by their natural product pathway, but by the structure of their substrate. (107) This relationship allowed the prediction of polyketide structure from the KS sequence alone. Structures of *trans*-AT KS domains with (108) and without (109) bound substrate highlight the single amino acid changes and large structural differences that give rise to the substrate specificity, which, in at least one case, was verified through site-specific

mutagenesis. (110) *In vitro* investigation of the psymberin and bacilaene *trans*-AT PKSs suggests that the observed specificity manifests at the elongation, not acylation, step of the reaction. (111)

Only recently was such a sequence-structure-selectivity relationship observed in a type I *cis*-AT PKSs. The giant aminopolyol synthases revealed a similar substrate-based clading for KS domains. (112) Zhang et al. identified conserved sequence motifs that correlate with the reductive state (hydroxyl, olefin, etc.) of the precursor. (112) There are several examples of substrate selectivity by *cis*-AT KS domains. Clear examples of substrate specificity are the KS domains that follow some iterative modules, termed “gatekeepers”. The related aureothin and neo-aureothin synthases contain iterative modules, AurA and NorB, that can extend their products between one and four times. In each case, however, the downstream modules only accept the second extension intermediate. (113, 114) Additionally, module 2 of the DEBS synthase was specific for some stereoisomers of its substrate. (115) Nevertheless, *cis*-AT KSs are not expected to be highly specific. Recently, KS substrate promiscuity was expanded by identification of an A154W mutation in the DEBS KS3 that conferred the ability to extend eight new substrates. (116) Robbins et al. identified conserved and variable residues in an alignment of 226 modular and 24 iterative KS domains, setting the stage for future mutagenesis studies (Figure 2-7). (117)

Once a target molecule has been identified, PKS modules with the desired activity can be combined in assembly lines. Computational tools, such as ClusterCad, can identify modules with the appropriate domains. (118) Rarely does a natively interacting pathway contain all the desired modules for the compound of interest. Fortunately, modules from different assembly lines can be engineered to interact with each other. Intermodular interaction is mediated by “linker domains” on the N- and C-termini of modules (Figure 2-7). (119) These small α -helical bundles can be installed on non-native modules to mediate interactions. In one study, of 154 unique bi-modular recombinations using this strategy, nearly half yielded detectable product. (120, 121) In another study, of 54 tri-modular recombinations, 96% yielded the expected product. (122)

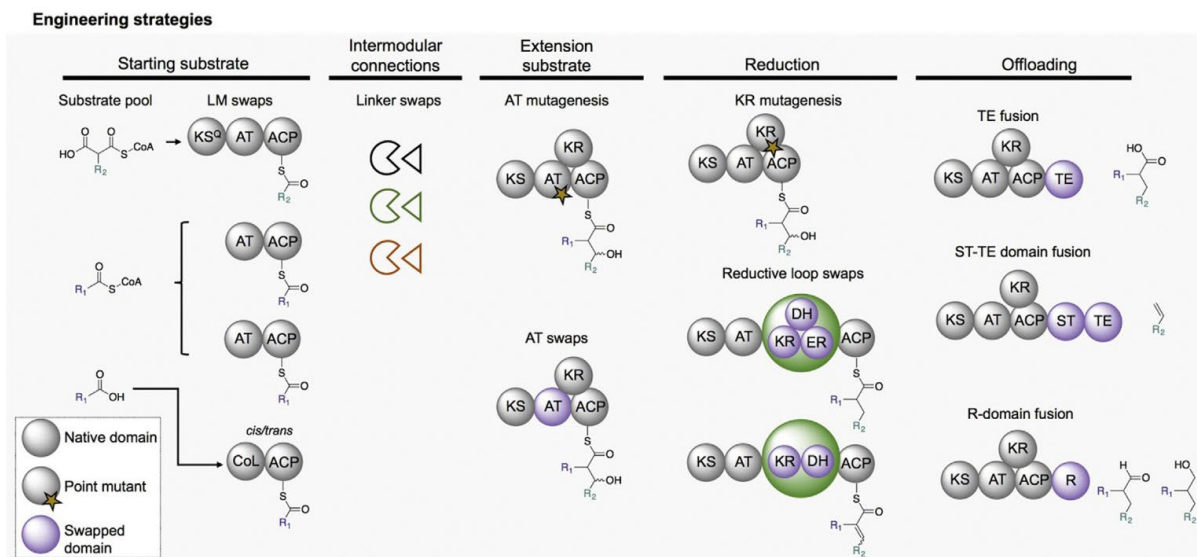


Figure 2-7. PKS engineering strategies for altering chain length and substitution.

2.3.2. α -Keto substitution

The α -keto substitution is primarily determined by the AT domain. Of 1129 PKSs analyzed, 55% of ATs were predicted to use malonyl-CoA while another 39% were predicted to use methylmalonyl-CoA. (123) Other extension substrates include ethylmalonyl-CoA, methoxymalonyl-CoA, and aminomalonyl-CoA. These are commonly produced through crotonyl-CoA carboxylase/reductase (CCR) pathways that cluster with the core biosynthetic pathway in the genome. (124)

Well before the first crystal KS-AT didomains were solved for DEBS M3 (125) and M5, (126) consensus sequences were discovered that correlate with malonyl-(HAFH) or methylmalonyl-CoA (YASH) specificity. (127) Mutation of the methylmalonyl-specific YASH motif to HAFH in DEBS AT1, (128) AT4, and AT6 (129) lead to loss of specificity for methylmalonyl-CoA. However, later in vitro investigations showed that these modules had greatly reduced activity. (130)

AT domain swapping is an additional technique for altering substrate specificity. Early domain swapping efforts relied on convenient restriction sites to construct chimeras, but nevertheless lead to the production of new compounds. For example, the DEBSM1-TE methylmalonyl-CoA specific AT was swapped for the rapamycin M2 malonyl-specific AT. (131) This led to the production of two novel desmethylated triketide lactones. PCR-based cloning and crystal structures of KS-AT fragments lead to more precise methods for making AT-swaps. (125, 126) Junction sites were identified with

minimal loss in turnover. (42) Preservation of structured regions in the post-AT linker was essential towards maintaining activity.

2.3.3. β -keto reduction

Several strategies exist for specifying the degree and stereochemistry of β -keto reduction. KR set most of the stereocenters in PKS modules and are the best-studied reducing domains. (132) KRs can be both stereospecific (binding one epimer over another to control α -substituent selectivity) and stereoselective (controlling which face of the β -keto group is attacked by the NADPH hydride to set the hydroxyl stereochemistry). (133) However, KRs have been shown to epimerize the α -substituent prior to reduction. (134) The A-type KRs generate a hydroxyl moiety with L-orientation and B-type KRs generate a hydroxyl moiety with D orientation (Figure 2-8. Stereochemical outcomes of β -keto reduction.). If the KR naturally acts on an α -substituted substrate, it can be further classified as A1 or B1 if the α -substituent is in the D-orientation or A2 or B2 if the α -substituent is in the L-orientation. (135) Structures are available for each type of KR (A1, (136) A2, (133) B1, (137) and B2 (138)) The presence of an LDD motif ~57 residues before the catalytic tyrosine indicates an B-type KR whereas its absence indicates an A-type KR. Additionally, the B-type KRs contain a tryptophan eight residues before the catalytic tyrosine. Structure-guided mutagenesis has led to the generation of A2-type (133) and non-stereospecific A-

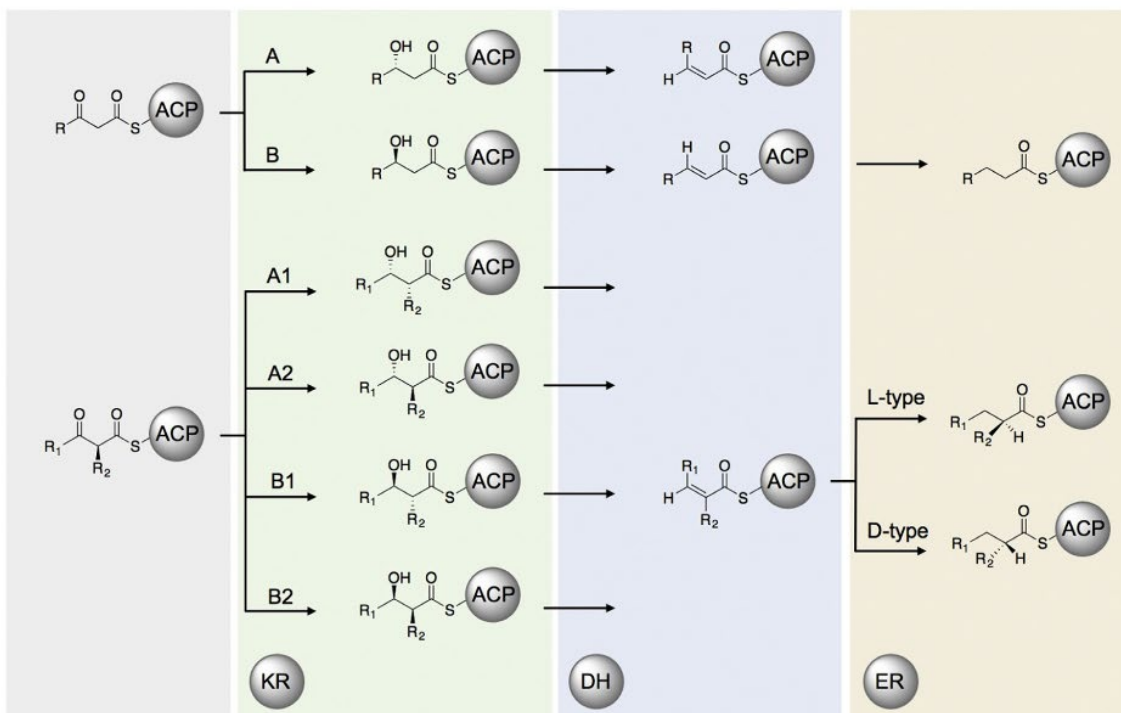


Figure 2-8. Stereochemical outcomes of β -keto reduction.

type KR (136) from A1-type KR. Additionally, a B2-type KR was mutated to an A2-type through a single point mutant. (139) However, Bailey et al. hypothesize that conversion to an A2 KR (anti addition) is energetically favored and converting from A2 to A1 or B2 (syn addition) will be more challenging. (139) Additionally, mutation of the catalytic tyrosine of the KR of the sixth module of the DEB synthase lead to production of a β -keto derivative, supporting that KR inactivation is a way to generate new polyketides. (140)

Dehydratase domains are responsible for installing α - β double bonds in PKSs. Dehydration requires the syn-coplanar elimination of water is therefore sensitive to the stereochemical centers of the substrate. (141) Thus, DH domains are tied to their preceding KR domains; most DHs follow B-type KR and install trans-double bonds.(135) However, both *cis*- and *trans*- double bonds are found in PKSs. Post-PKS processing is implicated in *cis*-double bond formation in borrelidin (142) and rifamycin (143), but *cis*-double bond of phoslactomycin is likely installed in situ by the first module of the synthase, Plm1. (144) Despite having several DH crystal structures, (143, 145, 146) there is no clear trend in the active site residues that govern the stereochemistry of reduction or the preference for α -substituents. (132) The specificity of DH domains past the α - and β - substituents is unknown.

Even less is known about the ER domains, which reduce trans- α - β -unsaturated intermediates provided by DHs. ERs can set the stereochemistry of the α -substituent in D- or L- orientation. (132) L-type ERs possess a conserved tyrosine residue that is absent in D-type KR. (147) Mutation of this residue in the ER from the fourth module of the DEB synthase switched the α -methyl substituent from S to R. (147) In the structure of Spn2, the ER from the second module of the spinosyn PKS this residue (Tyr421) was proximal to 4-pro-hydride of the bound NADP⁺, supporting its role as a proton donor. This structure also showed a lysine-aspartate pair that was crucial for catalysis. The lysine residue is within 5 Å of the tyrosine, and both are well positioned to donate a proton to the α -carbon of the substrate. In the absence of the tyrosine residue, the lysine may act as a proton donor. (148) This study also implicated a lysine-aspartate pair in catalysis and suggested that the lysine could donate a proton in the absence of tyrosine. The only structure available is the ER from the second module of the spinosyn PKS bound to NADP⁺. (148) This structure revealed that both the aforementioned tyrosine residue and a lysine residue may straddle the substrate. In the absence of the tyrosine, the lysine might donate a proton to the substrate.

In the absence of clear rules for reducing domain mutagenesis, domain or reducing loop swapping may be applied (Figure 2-7). Hagen et al. swapped the reductive loop from the first module of the borrelidin synthase for a domains from the spinosyn module B and borrelidin module 2 synthases, leading to the production of adipic acid. (96)

Reductive loop swaps have also been applied to generate analogs of erythromycin and ivermectin. (149–151)

2.3.4. Product release

The diversity of release mechanisms in type I PKSs has been reviewed extensively. (152) TEs are the best understood release domains, with several available crystal structures, known catalytic residues and mechanism. (135) Most TEs catalyze intramolecular cyclization to form macrolactones, but many catalyze intramolecular hydrolysis with water. (135) The substrate selectivity of type I PKS TEs has been reviewed extensively. (153) While TEs are generally selective in the chemistry they catalyze with their native substrates, such as the highly regio- (150) and stereospecific (154) cyclization by DEBS TE using the C13 nucleophilic alcohol, they are often versatile with non-native substrates. (153) This has been demonstrated with TE fusions (Figure 2-7). In one study, over 60 different macrolactones were released by DEB synthases with modified AT domains and reductive loops. (150) Additionally, non-native 6, (155) 8, (156) 12, (157) 14 and 16-membered rings (158) have been produced by either DEBS TE fusions or non-native substrate feeding. *In vitro* studies of isolated DEB (159, 160), pikromycin (161) and pimaricin (162) TE domains showed general hydrolysis of short non-native thioester substrates. However, the native substrates were not predictive of activity on non-native substrates. (160) The enzyme-substrate relationships that govern macrocyclization are not yet known.

Another attractive option for product release is reductive R domains (Figure 2-7). These NADPH-dependent domains catalyze 2- or 4-electron thioester reductions to aldehydes and alcohols, respectively. (152) Structure-guided mutagenesis of the terminal myxlamid R domain, MxaA, generated variants with increased activity towards highly-reduced substrates. (163) Finally, terminal alkene formation is possible through fusion of tandem sulfotransferase-TE domains (ST-TEs). These domains sulfonate the intermediate and then eliminate sulfate to produce an alkene (Figure 2-7). (164)

2.3.5. Persistent issues and the cutting edge of PKS engineering research

Two mechanisms underpin the success rate of PKS engineering. First, and perhaps lesser of the two issues, is substrate specificity of PKS modules. While PKSs are inherently promiscuous, (165) presentation of non-native substrates most often leads to detectable but greatly-decreased activity (99). Site-directed mutagenesis is one avenue to improve activity on non-native substrates and will be discussed below. Mounting evidence demonstrates that, in fact, improper protein-protein interactions are limiting. (166–168) (121, 122, 169) Module swapping requires the non-native interaction between head-to-tail neighbors at the ACP-KS interface. Structural and biochemical studies have

identified the residues that mediate these interactions, and mutation of the ACP can improve the efficiency of the interaction.(166–168) Similarly, domain swapping requires, at the very least, a non-native interaction between the ACP and the donor domain. Residues mediating AT-ACP and TE-ACP have been identified (Figure 2-9). (170, 171)

Gaining structural information about PKSs has been challenging due to the large size of PKS modules. However, the limited structural information has yielded engineering successes in leaps and bounds; KS-AT di-domain structures (125, 126) informed high-efficiency AT swaps (42) and linker domain structures identified the standard method for engineering intermodular interactions. Advances in cryo-electron microscopy permitted the first whole-module structure, pikAIII. Snapshots of the pikAIII module with substrates suggested major conformational changes at all stages of biocatalysis.(172, 173) The ACP domain of pikAIII synthase “swings” between active sites and is presented outside of the core domains after the appropriate elongation and reduction. Data from cryo-EM, SAXs and X-ray crystallography support the idea that FASs, PKSs, and NRPSs are all highly flexible and change conformations throughout their catalytic cycles.(174) Taking into account this structural information and the necessity to preserve conformational changes may increase our ability to engineer efficient non-native PKSs.

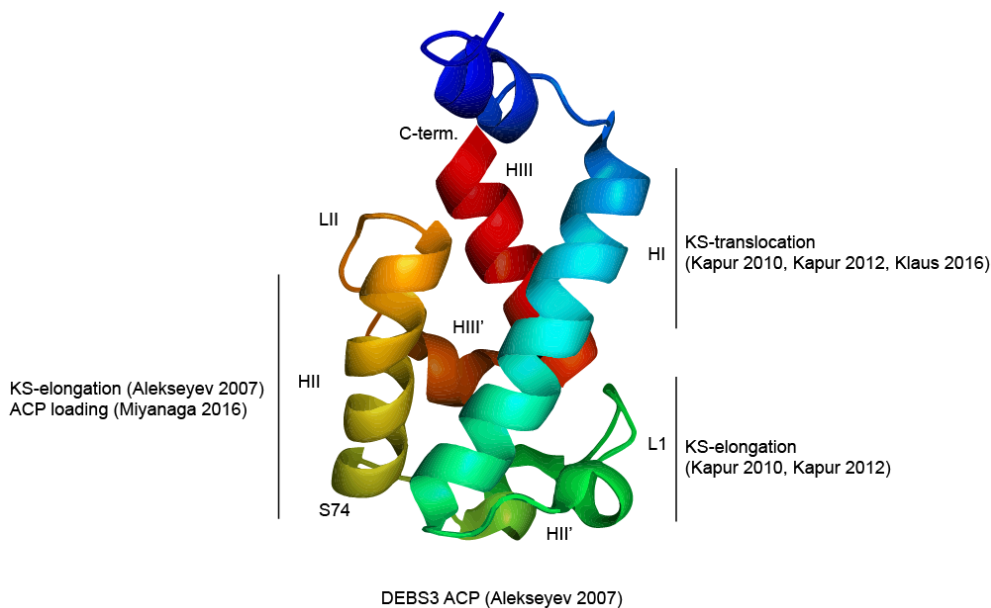


Figure 2-9. Structure of the DEBSM3-ACP. Helices mediating interactions are indicated with the original report.

Chapter 3. Probing the flexibility of an iterative modular polyketide synthase with non-native substrates in vitro

Contains material original co-authored as Curran S.C., Hagen A., Poust S., Chan L.J.G., Garabedian B.M., de Rond T., Baluyot M.J., Vu J.T., Lau A.K., Yuzawa S., Petzold C.J., Katz L., Keasling J.D. Probing the flexibility of an iterative modular polyketide synthase with non-native substrates in vitro. 2018. ACS Chemical Biology. 17; 13(8) 2261-2268. DOI: 10.1021/acscchembio.8b00422. Copyright © 2018 American Chemical Society. Reprinted with explicit co-author permissions and following publisher guidelines. The full text of this article can be accessed at <http://pubs.acs.org/articlesonrequest/AOR-g5VggqMagNQ8ejfjNeds>

3.1. Abstract

In the search for molecular machinery for custom biosynthesis of valuable compounds, the modular Type I polyketide synthases (PKSs) offer great potential. In this study we investigate the flexibility of BorM5, the iterative fifth module of the borrelidin synthase, with a panel of non-native priming substrates in vitro. BorM5 differentially extends various aliphatic and substituted substrates. Depending on substrate size and substitution BorM5 can exceed the three iterations it natively performs. To probe the effect of methyl branching on chain length regulation we engineered a BorM5 variant capable of incorporating methylmalonyl- and malonyl-CoA into its intermediates. Intermediate methylation did not affect overall chain length, indicating that the enzyme does not count methyl branches to specify the number of iterations. In addition to providing regulatory insight about BorM5, we produced dozens of novel methylated intermediates that might be used for production of various hydrocarbons or pharmaceuticals. These findings enable rational engineering and recombination of BorM5 and inform the study of other iterative modules.

3.2. Introduction

Our biosynthetic route towards methyl-branched methyl ketones (MBMKs) envisioned the use of the iterative module, BorM5. BorM5 is the fifth module Type I PKS that produces the nonaketide tRNA synthetase inhibitor, borrelidin (Figure 3-1). (175) BorM5 natively extends the highly-substituted pentaketide dicarboxylic acid intermediate three times with methylmalonyl-CoA before passing it to the downstream module, BorM8. The

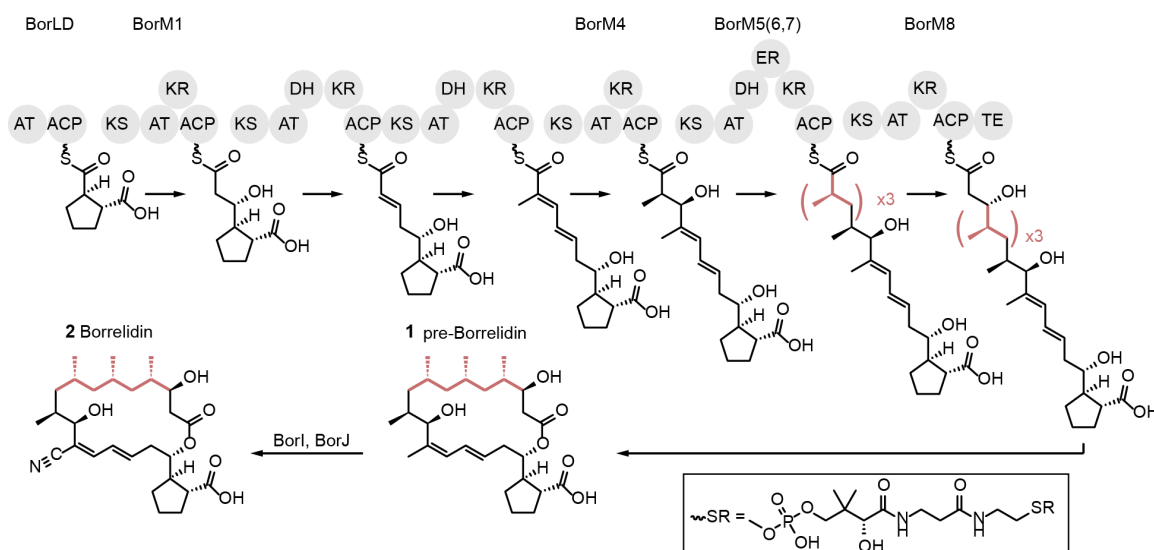


Figure 3-1. The borrelidin polyketide synthase pathway. This research was originally published as Olano C., et al. 2004. Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: cluster analysis and assignment of functions. *Chem Biol* 11(1):87–97. Copyright © Elsevier 2004. Reprinted with permission.

intermediate is then extended and cyclized to pre-borrelidin 1. (175) BorI and BorJ are non-PKS enzymes that install the carbonitrile moiety to produce mature borrelidin 2. (52) An alternative hypothesis to iteration is that three copies of the borrelidin gene product aligned to each extend the intermediate once. To verify that borrelidin was produced by a single BorM5 module and not identical modules arranged in series, BorM5 was fused to its up- and downstream neighbors. (176) This mutant produced borrelidin, confirming the iterative nature of BorM5.

The borrelidin synthase is not the only example of a modular Type I PKSs that does not follow co-linearity. In fact, many type I PKSs break the co-linearity paradigm. (123, 177) Aberrant iteration, or ‘stuttering’, was first observed in the otherwise colinear 6-deoxyerythronolide B (DEB) synthase, (178) followed by the epothilone (179) and pikromycin synthases. (180) For iteration to occur, the intermediate must be passed backwards after extension from the ACP to the KS of the same module, rather than to the KS of the downstream module. (181)

In the above cases, stuttering took place very infrequently, and did not significantly reduce the production levels of the natural product. “Programmed” iteration - wherein reuse of a module is required for the major product - was first observed in the stigmatellin synthase. (182) Additional examples of this family include the aureothin, (183) neo-aureothin, (184) borrelidin, (175) lankacidin, (185) etnangien,

(186) DKxanthene, (187) crocacin, (188) azalomycin F, (189) nocardiosis associated-(NOCAP), (190) and sceliphrolactam PKS. (191) These modular iterative PKSs (iPKSs) extend multiple, often highly substituted substrates in their native environments. This feature suggests that these enzymes may display enhanced promiscuity for non-native substrates as well, but this has not yet been investigated.

Chain length control is not well understood in these synthases. Studies of the related aureothin (AurA) and neo-aureothin (NorA) iterative modules and their respective assembly lines demonstrated that the downstream module, acts as a 'gatekeeper' and selects one of several intermediates from the stuttering modules; in its absence, the stuttering modules continued to extend past the correct length. (113, 114) Alternative hypotheses can be drawn from the related iterative fatty acid synthases (FAS) and other iterative PKSs, including Type II and fungal and bacterial iterative PKSs. *E. coli* keto-acyl-synthase II (KAS) from FAS controls chain length via a binding pocket in the KS. (192) The Type II iterative PKSs control chain length via an amphipathic binding tunnel in the KS-like chain length factor (CLF). (68, 193) Genetic recombination of KS α and CLF domains led to novel polyketides with varying lengths. (194) Studies with non-native priming substrates in the actinorhodin, (63) tetracenomycin, (64) benastatin, (65) and pradamycin (66) synthases revealed that the mechanism for chain length control is "measurement" of the polyketide chain. A crystal structure of the actinorhodin KS-CLF heterodimer revealed an amphipathic binding tunnel at the heterodimeric interface that limits chain length. (193) Amino acids impose "gates" on the chain length and their bulk is inversely correlated with chain length across multiple synthases. (68, 193) Mutation of these residues from large to small converted the actinorhodin synthase from an octaketide synthase to a decaaketide synthase. (68) Similarly, large-to-small mutations of these residues was sufficient to convert the decaaketide tetracenomycin synthase to an octaketide synthase. (68)

In fungal NR-PKSs, chain length control is likely mediated by the KS domain. (75) Several examples provide insight into the mechanism of chain length control. The PksA SAT-KS-MAT+ACP control was sufficient to produce an octaketide. (79) The truncated minimal PKSs containing the SAT-KS-MAT + ACP fragments from PksA and Pks4 (bikaverin synthase) produced the correct intermediates. (79, 195, 196) Trace amounts (<1%) of shorter intermediates were observed for Pks4. (195) It is not likely that the SAT or MAT domains influence chain length as they are solely implicated in primer and malonyl-CoA recruitment and attachment, respectively. A solution structure of the ACP domain from PksA did not suggest any involvement of the ACP in controlling chain length. (197) When Pks4 was primed with octanoyl-CoA it extended fewer times (from a nonaketide to a pentaketide synthase), suggesting it is indeed the actual length of the chain, and not the number of iterations that is important. In contrast, when the PR-PKS

6-methylsalicylic acid synthase was provided with non-native substrates in the absence of NADPH, it always extended twice. (198) This was interpreted as a potential “counting mechanism” but has not been elaborated upon. (72) The release of doubly-extended intermediates may simply reflect the inability of the synthase to stabilize poly- β -keto chain intermediates. For example, the HR-PKS nonaketide synthase LovB produced triketide lactone in the absence of NADPH. (199)

In this study, we leveraged the use of non-native substrates *in vitro* to explore the limits and regulation of iteration by BorM5. Purified BorM5 was reacted with a small library of non-native priming substrates *in vitro* and enzyme-bound intermediates were identified, providing insight into the intrinsic promiscuity of BorM5. We found that BorM5 iterates differentially depending on substrate size and substitution and can exceed the three iterations it catalyzes with the native substrate. We probed the effect of α -methylation on chain length determination by engineering BorM5 to accept the non-native extension substrate, malonyl-CoA. During this study, we identified several mechanisms for iteration regulation, produced novel methylated intermediates and demonstrated that BorM5 is amenable to domain engineering and is active with non-native substrates. Results and Discussion

3.2.1. Demonstrating BorM5 activity *in vitro*

BorM5 was expressed as a hexahistidine (6xHis) maltose binding protein (MBP) fusion to improve solubility and facilitate purification. (200) 6xHis-MBP-BorM5 and its upstream ACP, *apo*-BorACP4, were purified from *Escherichia coli* BL21 (DE3) *via* immobilized metal ion affinity and anion exchange chromatography (Supporting Figure 3-13). BorM5 and BorACP4 were phosphopantetheinylated *in vitro* to their corresponding *holo*- or acyl-forms, respectively, *via Bacillus subtilis* Sfp and acyl-CoAs (Supporting Figure 3-14). (201) The formation of acyl-BorACP4 was confirmed by the LC-MS/MS phosphopantetheine ejection assay.(202) Briefly, the ACP-bound phosphopantetheine arm (and acyl intermediate) is fragmented and releases a characteristic ion which provides additional confidence on the mass of the intermediate (Supporting Figure 3-15). Acyl-BorACP4 and *holo*-BorM5 were incubated with NADPH and methylmalonyl-CoA. In the absence of a TE domain, the intermediates remain bound to the BorM5 ACP and can be identified by the phosphopantetheine ejection assay (Figure 3-2).

We first assayed BorM5 with isobutyryl-BorACP4, as it mimics the α -substituent of the native substrates. As the productive concentration of acyl-BorACP4 and the reaction kinetics were unknown, we performed time course experiments with varying concentrations of isobutyryl-ACP4, ranging from 2 μ M to 200 μ M, or 0.5 x to 50 x molar equivalents of BorM5, and methylmalonyl-CoA and NADPH (Figure 3-3, panel b). After one hour, we observed ACP-bound intermediates whose abundances were dependent on

the concentration of isobutyryl-BorACP4. A representative chromatogram is shown (Figure 3-3, panel a). Low concentrations of isobutyryl-ACP4 (2 μ M and 4 μ M) produced small peaks corresponding to the second, third, and fourth extension intermediates (5, 6, 7) (Figure 3-3, panel b). These intermediates decreased in abundance over time, likely due to non-enzymatic hydrolysis. At 20 μ M isobutyryl-BorACP4, we observed second and third extensions after one hour at much greater abundance compared to the 2 and 4 μ M treatments. After 8 hours, we observed predominantly the third extension intermediate 6 (83% of the summed normalized peak areas for all extension intermediates (Figure 3-3) with minor amounts of extensions four (7) and five (8). At 24 hours, we observed a slight shift towards longer intermediates. For higher concentrations of ACP4 (40, 100, and 200 μ M), we also observed the doubly and triply extended intermediates 5 and 6 after one hour. However, the distributions remained relatively static; there was little conversion to longer species over time. No partially reduced intermediates were observed (Supporting Figure 3-16). Cleavage of the MBP-domain did not affect the distribution of intermediates, so it was left intact for remaining experiments (Supporting Figure 3-17).

Combined, as the concentration of isobutyryl-BorACP4 increased we observed 1) greater overall intermediate abundance, 2) relatively short intermediates due to slowed extension rates, and 3) an increase in the amount of isobutyryl-BorACP5 (3) (Figure 3-3, panel c). We propose that competition occurs between priming of the KS from an upstream ACP (isobutyryl-BorACP4) and iterative back-transfer (acyl-BorACP5), stalling BorM5 in the process of extending the ACP5-bound intermediate (Figure 3-3, panel d). The accumulation of isobutyryl-BorACP5 is due to non-enzymatic ACP-ACP

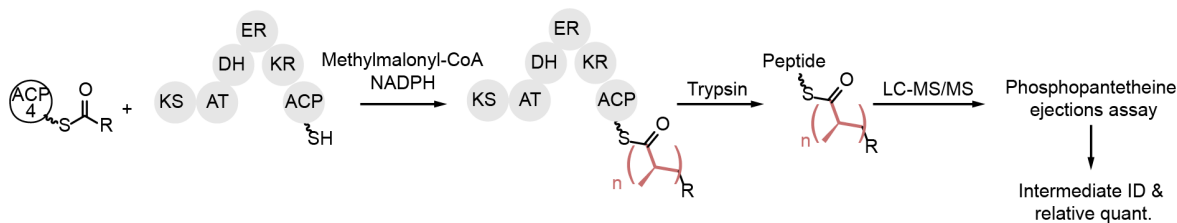


Figure 3-2. In vitro reaction setup and analysis.

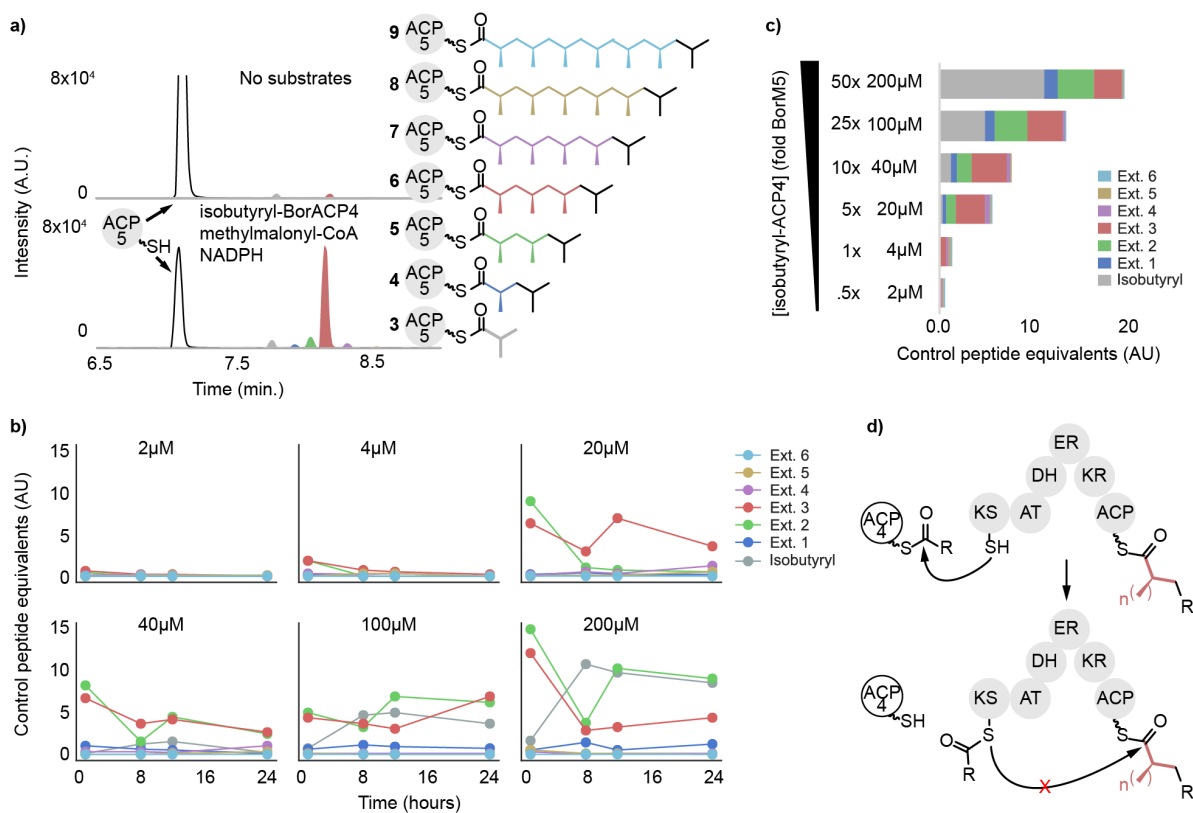


Figure 3-3. LC-MS/MS intermediate analysis of reactions with isobutyryl-ACP4, methylmalonyl-CoA, and NADPH. **a)** Representative LC-MS/MS traces at 8 hours for 4 μ M BorM5 with no substrates (top) or 20 μ M isobutyryl-ACP4 and methylmalonyl-CoA and NADPH (bottom). Color codes are constant throughout Figure 3 and correspond to extension iterations, shown to the right. Stereocenters are inferred from the structure of borrelidin, rather than experimental evidence. **b)** Time course assays of intermediates by LC-MS/MS. Each subplot corresponds to the concentration of isobutyryl-ACP4, as stated in the title. Time points are 1, 8, 12 and 24 hours. **c)** Summed intermediate distributions attached to BorM5 are displayed horizontally for different concentrations of isobutyryl-ACP4 after 8 hours. **d)** Scheme of proposed substrate inhibition by isobutyryl-ACP4. BorM5 is acylated at both the KS and ACP active sites (top) and is therefore unable to accept the ACP5-bound intermediate (bottom).

transthioesterification or KS-mediated transfer. The latter phenomenon, called module “skipping” has been observed in several systems. (203)

3.2.2. Expanding the library of non-native substrates

We next investigated BorM5’s promiscuity and chain length regulation with a panel of non-native substrates using equimolar BorM5: acyl-BorACP4 (10 μ M each). We prepared acyl-BorACP4 *via* Sfp-mediated phosphopantetheinylation of *apo*-BorACP4 (Supporting Figure 3-14). When BorM5 was primed with butyryl-BorACP4 **11**, we observed a distribution centered around the third extension intermediate **14** (62 \pm 17%) (Figure 3-4, panel a). When BorM5 was primed with hexanoyl-BorACP4 **12**, we observed a distribution centered around the 2nd extension intermediate **15** (67 \pm 21%). We did not observe intermediates when BorM5 was primed with decanoyl- or myristoyl-BorACP4, compared to the no-substrate control (see Raw Data). When BorM5 was primed with 3-hydroxybutyryl-BorACP4 **13**, which mimics the position of the hydroxyl in the native substrate, BorM5 strongly favored the five extensions, producing intermediate **16** (84 \pm 3.4%), significantly more iterations by BorM5 than with the native system.

To access a greater diversity of substrates and to determine if we saw the same extension characteristics as with acyl-BorACP4, we synthesized *S*-acyl-N-acetylcysteamine (SNAC) and *S*-acyl-thiophenol (TP) thioester analogues (See Chemical Synthesis). When BorM5 was primed with isobutyryl-SNAC **20**, again we observed the third extension **6** as the most abundant intermediate (48%) (Figure 3-5, panel a). Priming with acetyl-SNAC **18**, resulted in a distribution where the fourth extension intermediate **24** was the most abundant (73%). When BorM5 was primed with propionyl-SNAC **19**, we observed a distribution where the third extension intermediate **25** was the most abundant (87%). BorM5 accepted and extended 4-pentynoyl-SNAC **21** up to five times, while three iterations were favored (45%). Extension of *S*-(2*S*,3*R*)-3-hydroxy-2-methyl-SNAC (**22**), lead to primarily two but up to four extensions. To test if BorM5 could extend long diacids, as it does natively, we synthesized a 12-carbon diacid substrate, dodecanedioic-TP (See 3.4.4.). While we observed dodecanedioic-ACP5, we did not observe extension intermediates (See 3.4.9.). BorM5 extended hexanoyl-TP **23** up to three times, confirming that the TP handle was accepted (Figure 3-5, panel a).

To estimate a rough yield for each of the priming substrates used, we compared the loss of *holo*-BorACP5 between treatments after 30 minutes of elongation (Supporting Figure 3-18). *holo*-ACP5 and acyl-ACP5 are inversely correlated. As the concentration of BorM5 is known from the beginning, comparison of *holo*-BorM5 and reacted BorM5 may give the ACP acylation yield in molar units. This semi-quantitative approach should be considered carefully, however, as acylation by any acyl group (extension substrate or priming substrate) may be erroneously perceived as activity. We observed that methylmalonyl-CoA accounted for ~25% of the loss of the *holo*-ACP (1.25 μ M). The acyl-ACP4 substrates did not lead to greater acylation, despite the appearance of their extension intermediates. This is likely due to suboptimal concentrations. Incubation with acyl-SNACs and methylmalonyl-CoA lead to the loss of 50–70% of *holo*-BorACP5. Acetyl-SNAC was the least-preferred substrate, while isobutyryl- and pentynyl-SNAC were the most preferred.

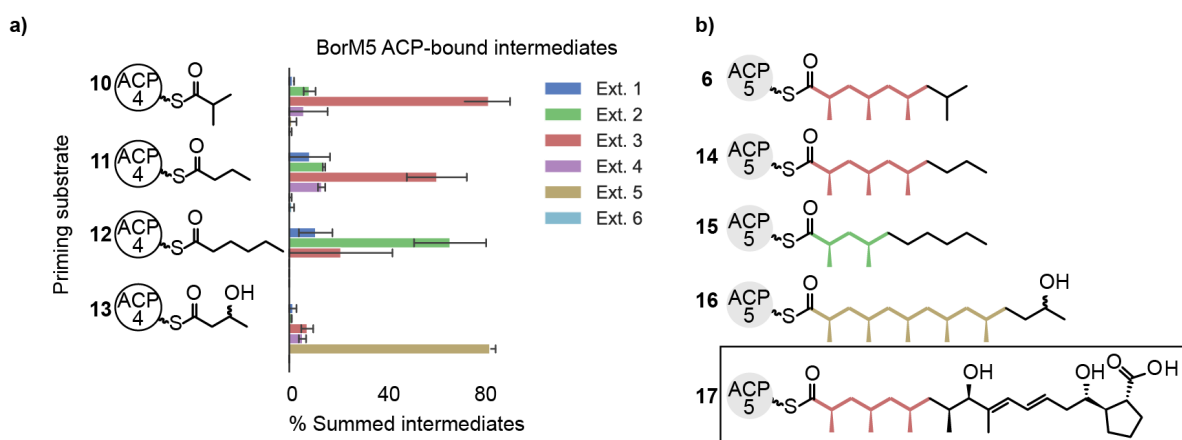


Figure 3-4. LC-MS/MS analysis of BorM5-bound intermediates after 8 hours of extension with acyl-BorACP4, methylmalonyl-CoA, and NADPH. **a)** Intermediates observed for each priming substrate (left) are depicted in proportion to the sum of all extended intermediate peak area. Error bars are standard deviation of ≥ 2 experiments. **b)** The most abundant intermediates for each priming substrate are shown. Intermediates are colored by the number of extensions, while the priming unit is in black. The native 3-extension intermediate is shown below in the box for comparison (not experimentally derived). Stereocenters are inferred from the structure of borrelidin.

3.2.3. Engineering BorM5 to accept malonyl-CoA.

The intermediates observed throughout our study did not reach their expected chain lengths given the native product **17** (Figure 3-4 panel b, Figure 3-5, panel b). Rather, the major intermediates mostly corresponded to two and three extensions, except for 3-hydroxybutyryl-CoA which was extended five times. We reasoned that the methyl groups installed at each round of extension might be recognized by BorM5 and could contribute to chain length regulation, however loosely. The AT domain is responsible for selecting the extension substrate (*e.g.*, methylmalonyl-CoA) and therefore the presence of methyl substituents on the extended intermediates. Using a set of design guidelines recently published for homology-based AT domain swapping, (42) we constructed a variant of BorM5 that contained the AT domain from module 4 of the epothilone synthase, which accepts both methylmalonyl- and malonyl-CoA. (204) After one hour, this mutant, BorM5-epoAT, extended isobutyryl-SNAC up to four times with methylmalonyl-CoA (Figure 3-6, panel a), comparable to wild type (WT) BorM5. When BorM5-epoAT was incubated with malonyl-CoA and isobutyryl-SNAC, we observed up to four extensions (**29–32**), which were not observed when using WT BorM5. The second extension was the most abundant. Finally, BorM5-epoAT incubated with a 1:1 mixture of malonyl- and methylmalonyl-CoA yielded many species ranging from C6 to C13 (Figure

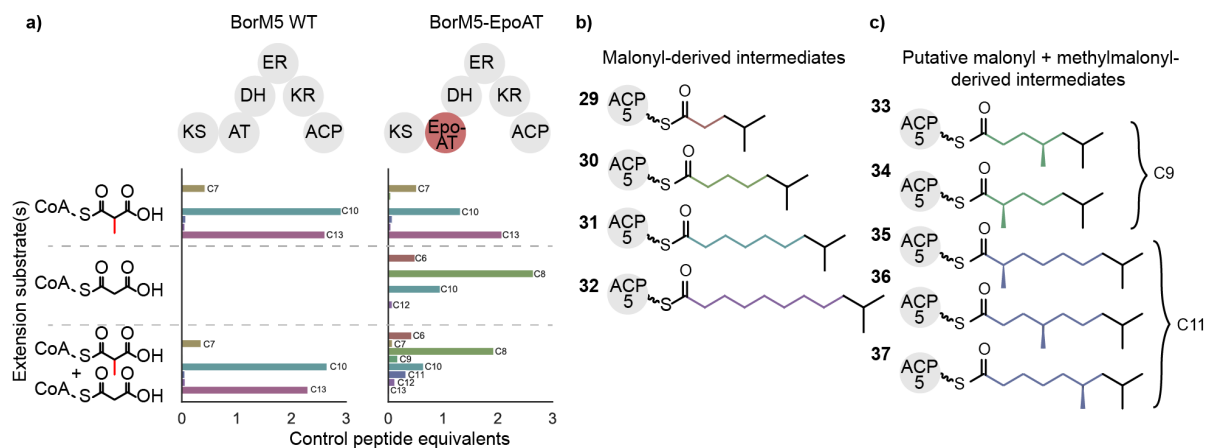


Figure 3-6. LC-MS/MS analysis of assays of BorM5 WT and BorM5-EpoAT-bound intermediates after 3 hours of extension with isobutyryl-SNAC, malonyl- and/or methylmalonyl-CoA, and NADPH. **a)** LC-MS/MS intermediate analysis of BorM5 WT (left) or Epo-AT (right) with isobutyryl-SNAC and either malonyl- or methylmalonyl-CoA. N=1. **b)** Intermediates derived from extension with malonyl-CoA and **c)** putative intermediates derived from methylmalonyl- and malonyl-CoA. Stereocenters are inferred from the structure of borrelidin, rather than experimental evidence.

3-6, panel b, c). Most of these likely correspond to the incorporation of malonyl-CoA, but the C9 (33 and 34) and C11 (35–37) species correspond to the incorporation of both malonyl- and methylmalonyl-CoA, although it is not possible to tell which isomer is produced from these data.

3.3. Conclusion

Earlier *in vivo* heterologous expression studies indicated that the borrelidin cluster was highly specific. No congeners of borrelidin--corresponding to fewer or more iterative extensions--were observed during heterologous expression of the *bor* cluster. (175) When the borrelidin cluster was provided with an array of non-native substrates *in vivo*, only diacid substrates were incorporated into borrelidin analogues. (104) This is likely due to the specificity of the loading module, BorLD, which is selective for diacid substrates *in vitro*. (105) Isolated BorM5, on the other hand, shows great flexibility for non-native substrates. We observed enzyme-bound intermediates ranging from one to six extensions of the priming substrates. Shorter intermediates were present after one hour, and indeed within one minute with isobutyryl-SNAC (Supporting Figure 3-19). We observed these long intermediates after 8 hours, well beyond what is expected of a canonical PKS. For example, the reconstituted hexamodular DEB synthase completes approximately one turnover per minute. (205) Nevertheless, the observation that BorM5 can extend past its native three extensions is in line with previous studies of the iterative modules of the aureothin and neo-aureothin synthases, which could also exceed the number of iterations required for their primary products. (113, 114)

The structure of the starting substrate leads to a different number of iterations, which is in line with studies of Type II PKSs and fungal iterative PKSs. (68, 195) From the intermediates observed here, it is challenging to determine an overall pattern of extension. We observed an inverse correlation between the length of aliphatic starting substrates and the number of extensions, but these intermediates did not reach the length of the native substrate (Figure 3-4 panel b, Figure 3-5, panel b). 3-Hydroxybutyryl was extended five times, compared to three times for butyryl at a similar time point, suggesting that BorM5 may require specific protein-substrate contacts for later extensions.

BorM5 natively extends with methylmalonyl-CoA and its intermediates are therefore α -methyl-branched. These branches could be used to enforce a “counting mechanism” that limits iteration to three times, regardless of starting substrate. To test this hypothesis, we generated an AT-swapped variant of BorM5 that incorporates malonyl-CoA. The intermediates extended with malonyl-CoA do not reach greater lengths, as would be expected if the methyl branches were limiting (Figure 3-6).

Which domain controls iteration? We can exclude the AT, as swapping the Epo-AT for the BorM5 AT did not significantly affect iteration. As we did not observe any partially reduced intermediates with isobutyryl-ACP (we did not examine the presence of partially reduced intermediates for other starting substrates), it appears that the reductive domains do not control iteration. This leaves the KS domain as the likely candidate for controlling iteration, like the previously discussed iterative systems. Structural studies will greatly aid our understanding of substrate recognition by BorM5.

In all cases where we observed intermediates, we observed a distribution of intermediates. The iterative/modular aureothin module AurA also produces such a distribution, ranging from 1 to 4 extensions, when heterologously expressed by itself *in vivo*. (113) In our case, we attribute this observation, at least partially to the interruption of iteration by re-loading of a new priming unit. In order for iteration to occur, the intermediate must pass backwards from the ACP to the KS of the same module. (181) If the KS active site cysteine is acylated, iteration should stall until the KS is free to perform nucleophilic attack (Figure 3-3, panel **d**). The related mammalian FAS is subject to substrate inhibition at high levels of its priming substrate, acetyl-CoA. (206) Similarly, as the concentration of the priming unit increased, BorM5 produced shorter intermediates (Figure 3-3, panel **c**). It may be for this reason that the *bor* cluster contains a Type II *trans*-thioesterase (TEII), BorB, (175) which might remove aberrant or stalled intermediates from BorM5 ACP. (207) An alternative explanation for a distribution is the possibility that the half-life of BorM5 *in vitro* is such that it expires during the course of the assays. However, BorM5 was equally active after 12 hours of incubation at 25°C (Supporting Figure 3-20). Alternatively, the substrate pool might be exhausted due to activity of co-purified protein contaminants or spontaneous oxidation or hydrolysis before BorM5 can fully extend its intermediates. However, re-addition of substrates after 12 hours did not alter the distribution of intermediates, arguing against this possibility (Supporting Figure 3-21).

These data favor kinetic control as the driver of fidelity in borrelidin biosynthesis. In this model, kinetic competition occurs between the domains rather than the execution of a programmed sequence of operations. The acyl-ACP performs a “random walk” between domains of various specificity for each intermediate. Such a phenomenon was observed for the highly regioselective methyltransferase domain of the fungal iterative synthase, LovB. (208) α -methyl transfer occurs only on one of the eight iterations. The optimal timing of this reaction is managed not by programmed sequential visits to each active site, but by higher kinetic activity for the correct substrate by the methyltransferase. β -keto reduction, on the other hand, is applied during each extension and the KR is relatively substrate agnostic. As for the borrelidin synthase, kinetic competition may occur between the KS domains of BorM5 and the downstream module, BorM8. As the

chain length grows, BorM5 may slow and permit downstream transfer, as observed for isobutyryl-ACP4 in this study (Figure 3-3, panel b). BorM8 may be a highly specific gatekeeper in order to compensate for the promiscuity of BorM5, as was shown for the aureothin and neo-aureothin synthases. (113, 114)

In conclusion, the *in vitro* approaches used in this study more clearly define the regulation of iteration by BorM5. By varying the priming substrate, we observed intermediates of different chain lengths and number of ketide units. On the other hand, varying the extension substrate from methylmalonyl- to malonyl-CoA only weakly affected the intermediate distributions. BorM5 can exceed the number of iterations it catalyzes on its native substrate, arguing against a counting mechanism for controlling iteration. Throughout these assays, BorM5 generated dozens of novel multi-methyl-branched intermediates, demonstrating its flexibility. Recently a fungal iterative PKS was shown to extend over 25 non-native substrates, suggesting iterative PKSs may be generally promiscuous. (209) This study opens the door for using BorM5 in an engineered assembly line.

3.4. Materials and methods

3.4.1. Reagents

All reagents were purchased from Sigma unless otherwise noted.

3.4.2. Cloning

Plasmids used in this study are summarized (Supporting Table 3-1). Cloning was facilitated with the J5 DNA assembly software. (210)

Supporting Table 3-1. Plasmids used.

Name	Backbone	Insert	Tag	Source	Accession
pARH130	pBBS2K	BorM5	N-term. 6xHis- MBP-TEV	This study	https://registry.jbei.org/entry/45070
pSC005	pBBS2K	BorM5- epoAT	N-term. 6xHis- MBP-TEV	This study	https://registry.jbei.org/entry/78786
pARH167	pET28a	BorACP 4	N-term. 6xHis- TEV	This study	https://registry.jbei.org/entry/45193
pRK793	pMal-C2	TEV protease	N-term. 7xHis- MBP-TEV	David Waugh (211)	https://www.addgene.org/8827/
pET-SFP	pET	SFP	N-term. 6xHis	Jacquelyn Blake- Hedges	https://registry.jbei.org/entry/58169

pARH130(6xHis-MBP-TEV-BorM5): The coding region for borrelidin module 5 (bora5; Uniprot accession #Q70HZ8)3 was codon optimized for *E. coli* and synthesized by IDT. Bora5 was amplified by Q5 polymerase (NEB), digested by NheI/XhoI (Thermo FastDigest) and ligated into similarly digested pBBS2K-6xHis-MBP-TEV.(212) DH10B transformants were sequence verified.

pSC005(6xHis-MBP-TEV-BorM5-EpoAT): Conserved KS-AT linker regions for BorM5 and epothilone module 4 AT (N-terminal junction: ...HVIVVEEAPAV..... GRRVPLPTY... Colors: BorM5, EpoM4): were determined based on homology. (42) Primers were designed with Device Editor. (213) DNA fragments were amplified by Q5 polymerase (NEB) and assembled via Gibson assembly (NEB). (214) DH10B transformants were sequence verified.

pARH167(6xHis-BorA4-ACP): The coding region for borrelidin module 4 ACP3 (bora4_ACP; Uniprot accession # Q70HZ9) and N- and C-terminal linkers were identified by MAPSI. (215) Bora4_ACP was amplified by Q5 polymerase (NEB), digested by NheI/XhoI (Thermo FastDigest) and ligated into similarly digested pET28a (Novagen). DH10b transformants were sequence verified.

3.4.3. Acyl-CoAs

All acyl-CoAs were purchased from Sigma.

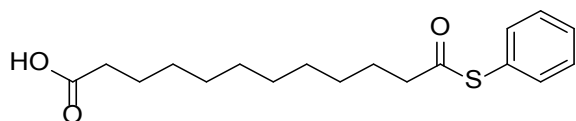
3.4.4. Acyl-SNAC/TP synthesis

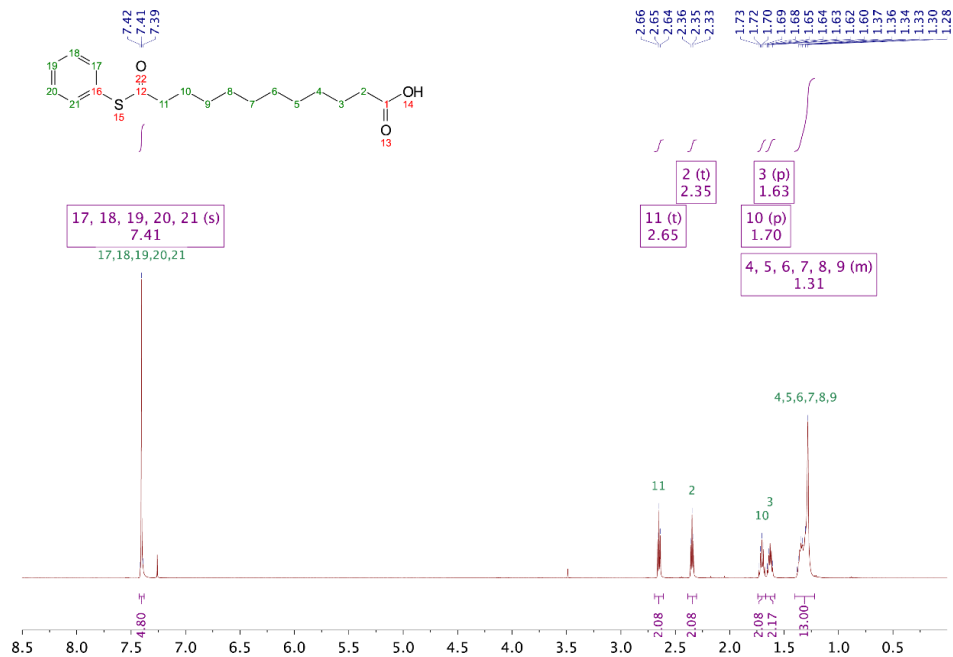
12-oxo-12-(phenylsulfanyl)dodecanoic acid

To a flame-dried 10 ml round-bottom flask was added 1mmol (230 mg) dodecanedioic acid dissolved in 1ml anhydrous tetrahydrofuran (THF). The flask was flushed with nitrogen and 1.1mmol (170 μ l) diisopropylcarbodiimide (DIC) in 500 μ l anhydrous THF was added and allowed to react for 10 minutes at ambient temperature with stirring. Next, 1.1mmol (112 μ l) thiophenol in 500 μ l anhydrous THF was added drop-wise over a 1h period and allowed to react for an additional 3h at ambient temperature. The product mixture was syringe-filtered (0.2 μ m PTFE Acrodisc CR13) and the compound purified from the reaction crude by flash chromatography (ethyl acetate:hexanes, 1:3). The pure product was concentrated under reduced pressure to give the title compound as a white crystalline solid (234mg, 73% yield).

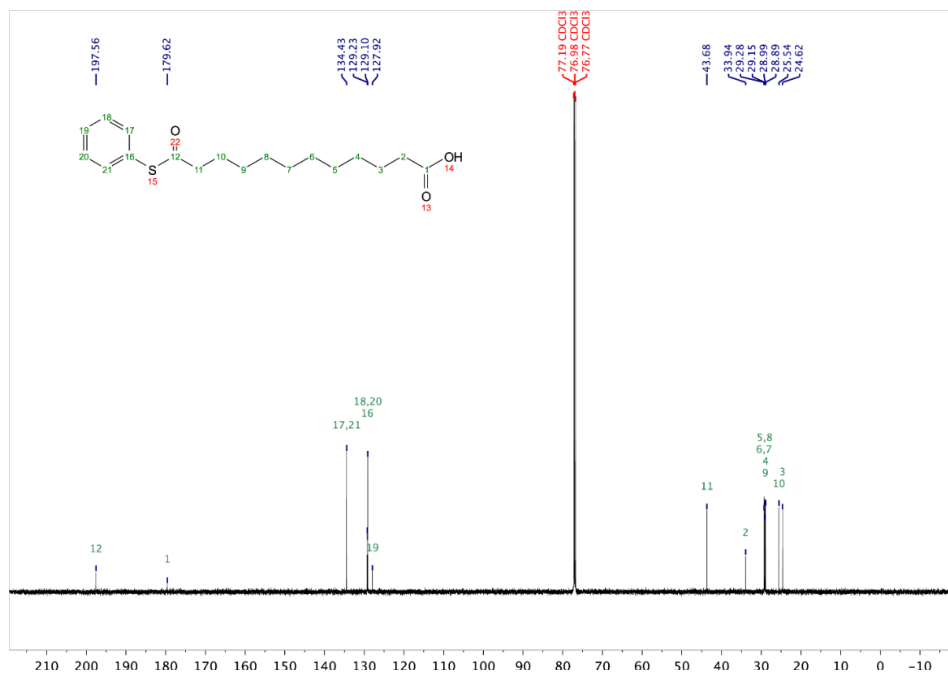
^1H NMR (600 MHz, Chloroform-*d*) δ 7.41 (s, 5H), 2.65 (t, J = 7.5 Hz, 2H), 2.35 (t, J = 7.5 Hz, 2H), 1.70 (p, J = 7.5 Hz, 2H), 1.63 (p, J = 7.5 Hz, 2H), 1.40 – 1.22 (m, 12H).

^{13}C NMR (151 MHz, Chloroform-*d*) δ 197.56, 179.62, 134.43, 129.23, 129.10, 127.92, 43.68, 33.94, 29.28, 29.15, 28.99, 28.89, 25.54, 24.62.





Supporting Figure 3-1. ¹H NMR spectrum of 12-oxo-12-(phenylsulfonyl)dodecanoic acid. CDCl₃. 600 MHz.



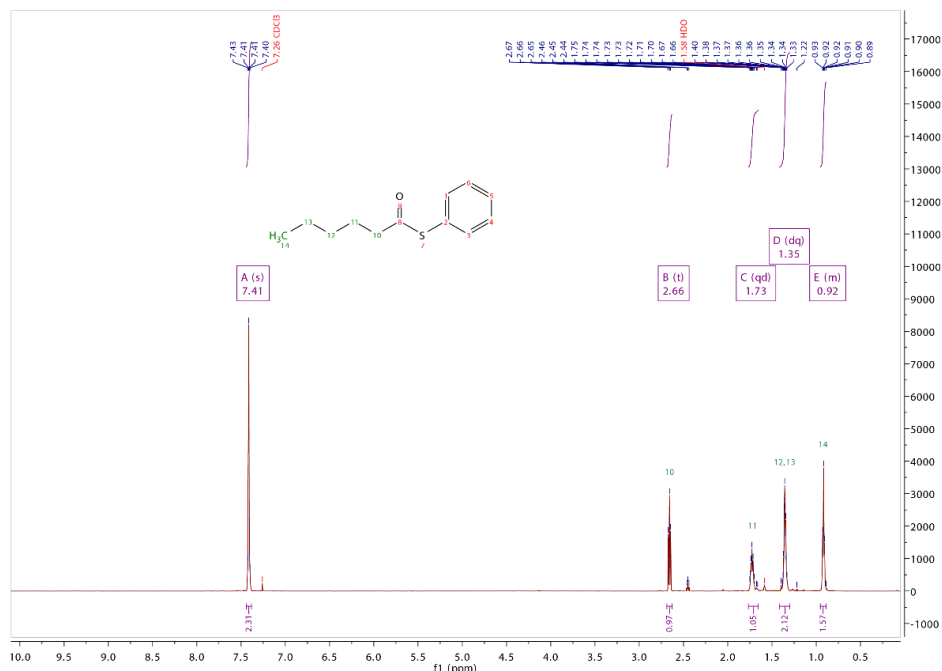
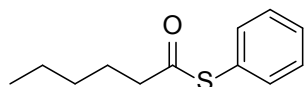
Supporting Figure 3-2. ¹³C NMR spectrum of 12-oxo-12-(phenylsulfonyl)dodecanoic acid. CDCl₃. 151 MHz.

S-phenyl hexanethioate

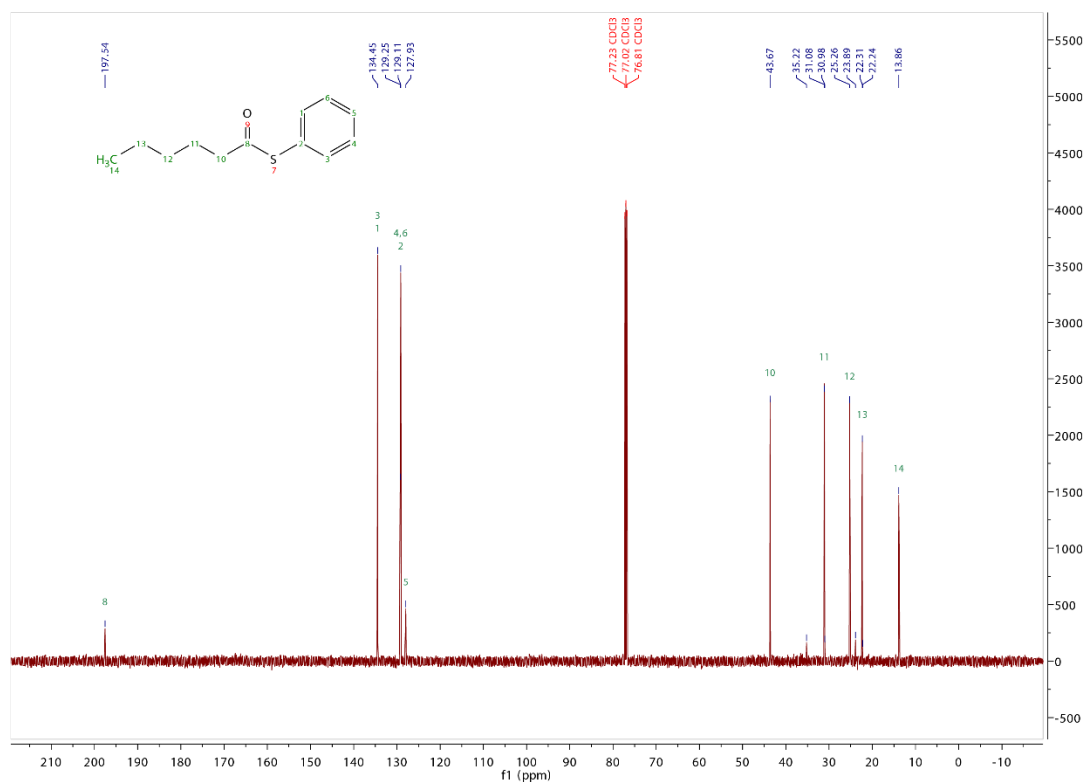
To a flame-dried 10 ml round-bottom flask was added 1mmol (125 μ l) hexanoic acid. The flask was flushed with nitrogen and 1.1mmol (170 μ l) DIC in 500 μ l anhydrous THF was added and allowed to react for 5 minutes at ambient temperature with stirring. Next, 1.1mmol (112 μ l) thiophenol in 500 μ l anhydrous THF was added drop-wise over a 1h period and allowed to react for an additional 3h at ambient temperature. The product mixture was syringe-filtered and passed over a bed of CuSO₄-impregnated silica to remove unreacted thiols. The product was purified by flash chromatography (hexanes) and concentrated under reduced pressure to afford the title compound as a transparent colorless oil (178mg, 86% yield).

¹H NMR (600 MHz, Chloroform-*d*) δ 7.41 (s, 5H), 2.66 (t, *J* = 7.5 Hz, 2H), 1.73 (p, *J* = 7.4 Hz, 2H), 1.39 – 1.31 (m, 4H), 0.94 – 0.89 (m, 3H).

¹³C NMR (151 MHz, Chloroform-*d*) δ 197.55, 134.45, 129.25, 129.11, 127.93, 43.67, 31.08, 25.26, 22.30, 13.86.



Supporting Figure 3-3. ¹H NMR spectrum of S-phenyl hexanethioate. CDCl₃. 600 mHz.



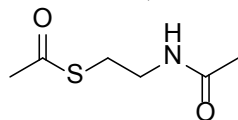
Supporting Figure 3-4. ¹³C NMR spectrum of S-phenyl hexanethioate. CDCl₃. 151MHz.

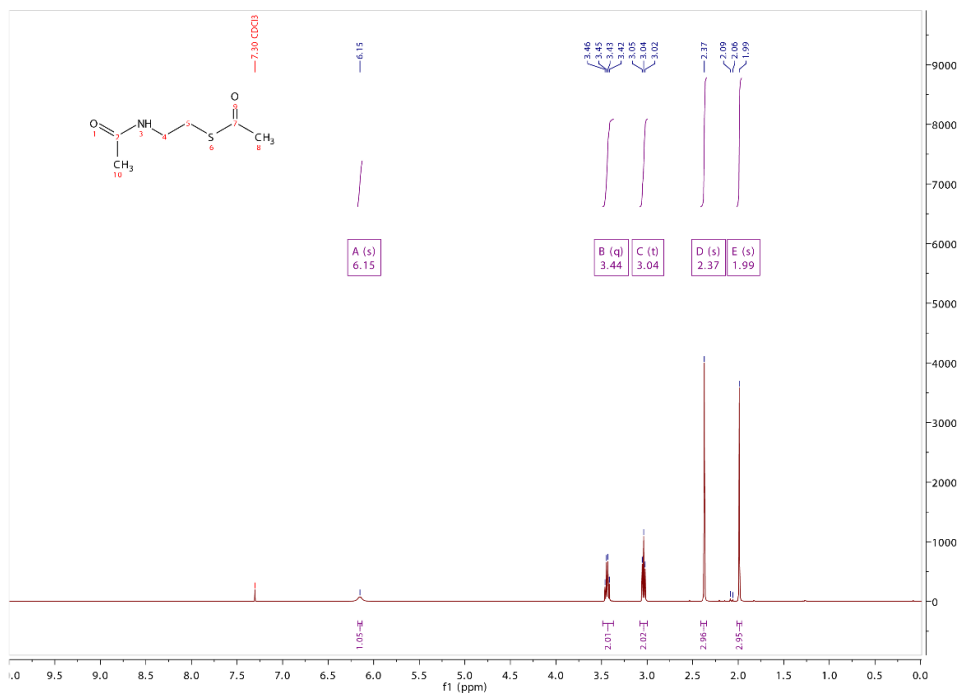
S-Acetyl-N-acetylcysteamine

To a glass vial, 1.7mmol (105 mg) acetic acid, 3.5 mmol (670 mg) 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) HCl, and 0.17mmol (21 mg) 4-Dimethylaminopyridine (DMAP) was dissolved in 2.5 mL dichloromethane (DCM) and allowed to react for 10 minutes at ambient temperature while stirring. Next, 2.1mmol (222.5μl) acetylcysteamine (HSNAC) in 2.5 mL DCM was added to the reaction mixture and allowed to react overnight at ambient temperature. The product was syringe-filtered and extracted with water. The compound was purified by silica column chromatography (ethyl acetate:hexanes, 0-100% ethyl acetate over 20CV). Pure product was concentrated under a rotary evaporator to yield the title compound as a solid (39 mg, 13.8%).

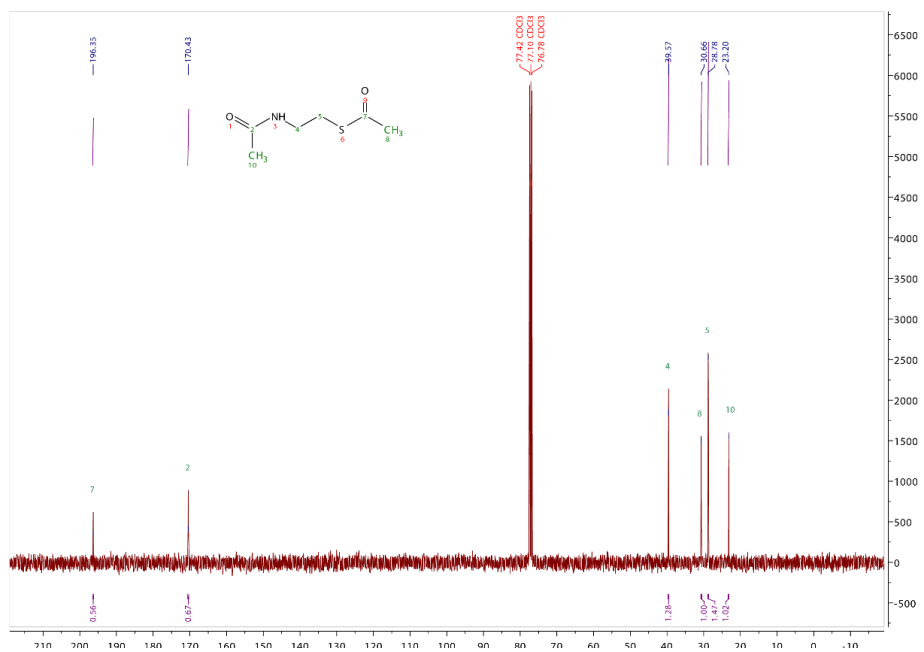
¹H NMR (400 MHz, Chloroform-*d*) δ 6.15 (s, 1H), 3.44 (q, *J* = 6.3 Hz, 2H), 3.04 (t, *J* = 6.5 Hz, 2H), 2.37 (s, 3H), 1.99 (s, 3H).

¹³C NMR (101 MHz, Chloroform-*d*) δ 196.35, 170.43, 39.57, 30.66, 28.78, 23.20.





Supporting Figure 3-5. ¹H NMR spectrum of acetyl-N-acetylcysteamine. CDCl₃. 400 mHz.



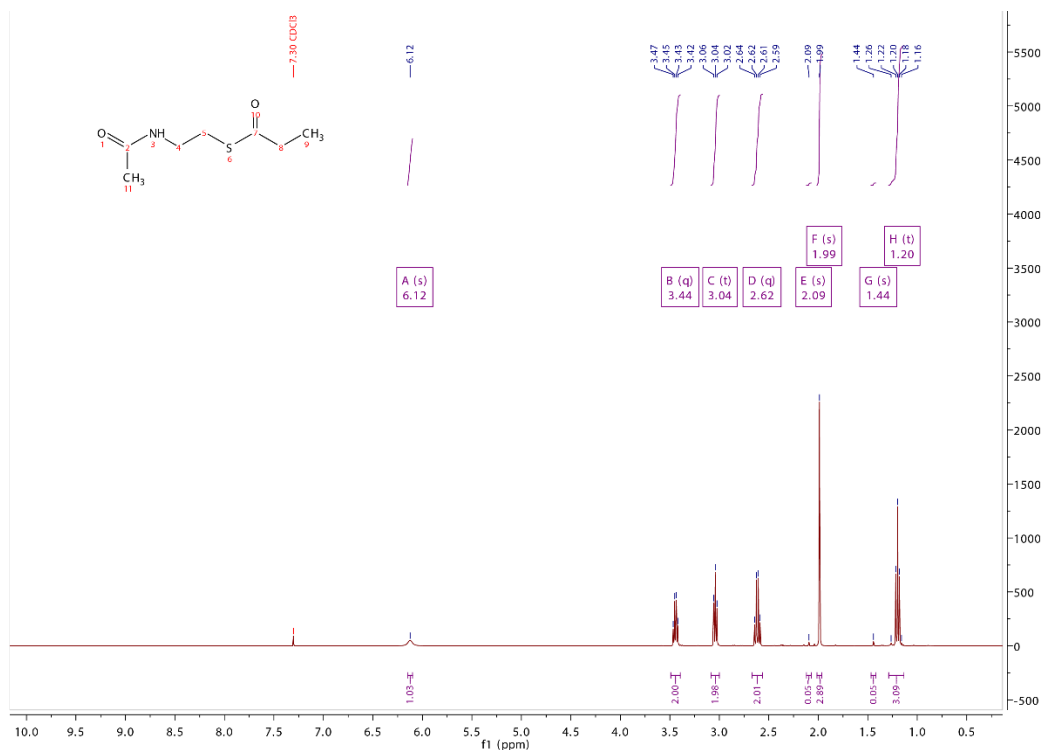
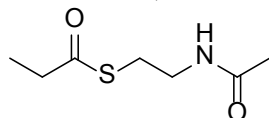
Supporting Figure 3-6. ¹³C NMR spectrum of acetyl-N-acetylcysteamine. CDCl₃. 151MHz

S-Propionyl-N-acetylcysteamine

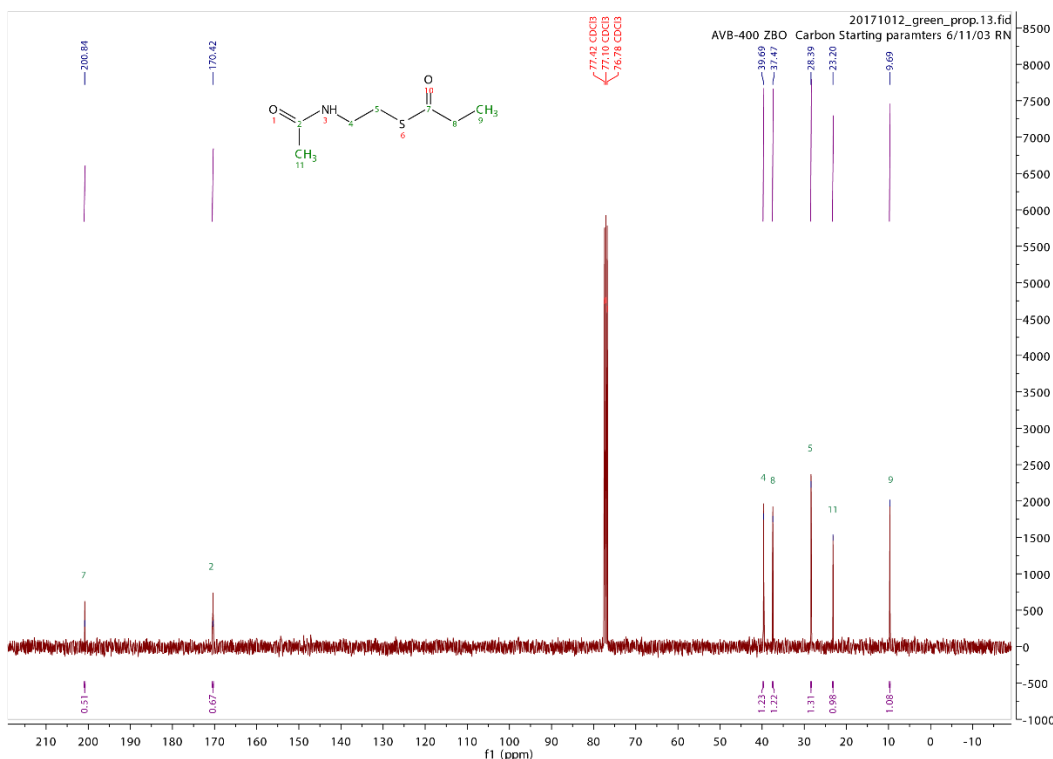
To a glass vial, 1.3mmol (99mg) propionic acid, 2.6mmol (510 mg) EDC HCl, and 0.13mmol (20 mg) DMAP was dissolved in 2.5 mL DCM and allowed to react for 10 minutes at ambient temperature while stirring. Next, 1.6mmol (169.9 μ l) acetylcysteamine (HSNAC) in 2.5 mL DCM was added to the reaction mixture and allowed to react overnight at ambient temperature. The product was syringe-filtered and extracted with water. The compound was purified by silica column chromatography (ethyl acetate:hexanes, 0-100% ethyl acetate over 20CV). Pure product was concentrated under a rotary evaporator to yield the title compound as a solid (39 mg, 16%).

^1H NMR (400 MHz, Chloroform-*d*) δ 6.12 (s, 1H), 3.44 (q, $J = 6.2$ Hz, 2H), 3.04 (t, $J = 6.5$ Hz, 2H), 2.62 (q, $J = 7.5$ Hz, 2H), 1.99 (s, 3H), 1.20 (t, $J = 7.5$ Hz, 3H).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 200.84, 170.42, 39.69, 37.47, 28.39, 23.20, 9.69.



Supporting Figure 3-7. ^1H NMR spectrum of propionyl-N-acetylcysteamine. CDCl_3 , 400 MHz.



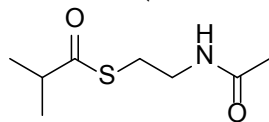
Supporting Figure 3-8. ^{13}C NMR spectrum of propionyl -N-acetylcysteamine. CDCl_3 . 101MHz.

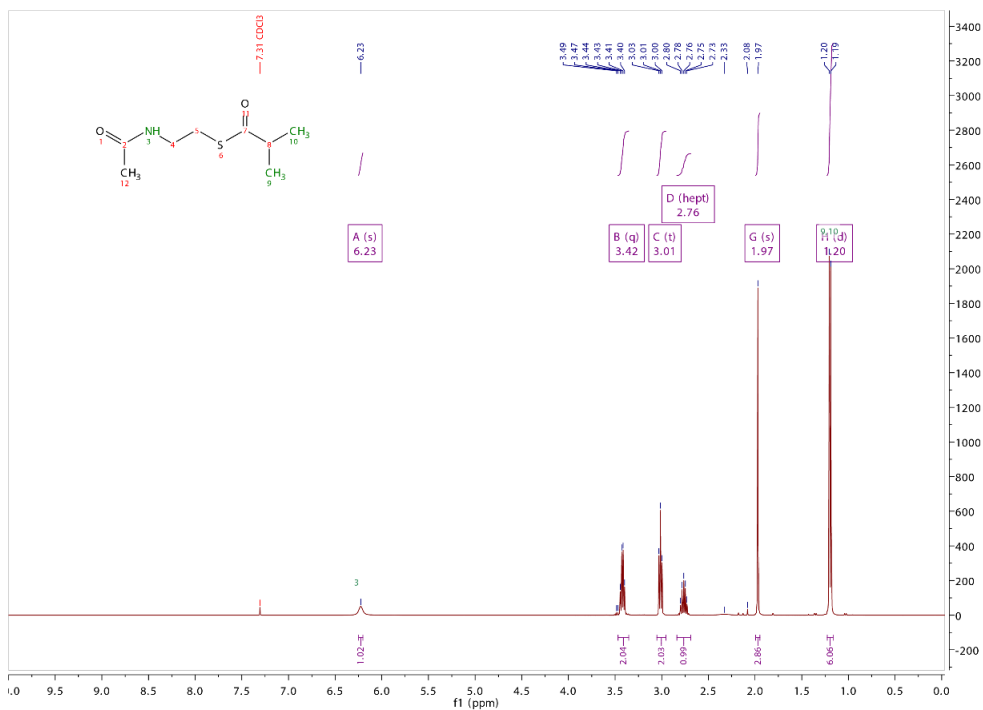
S-Isobutyryl-N-acetylcysteamine

To a glass vial, 1.7mmol (153.9mg) isobutyric acid, 3.49mmol (670 mg) EDC HCl, and 0.175 mmol (20 mg) DMAP was dissolved in 2.5 mL DCM and allowed to react for 10 minutes at ambient temperature while stirring. Next, 2.1mmol (222.5 μl) acetylcysteamine (HSNAC) in 2.5 mL DCM was added to the reaction mixture and allowed to react overnight at ambient temperature. The product was syringe-filtered and extracted with water. The compound was purified by silica column chromatography (ethyl acetate:hexanes, 0-100% ethyl acetate over 20CV). Pure product was concentrated under a rotary evaporator to yield the title compound as a solid (49.5 mg, 13.8%).

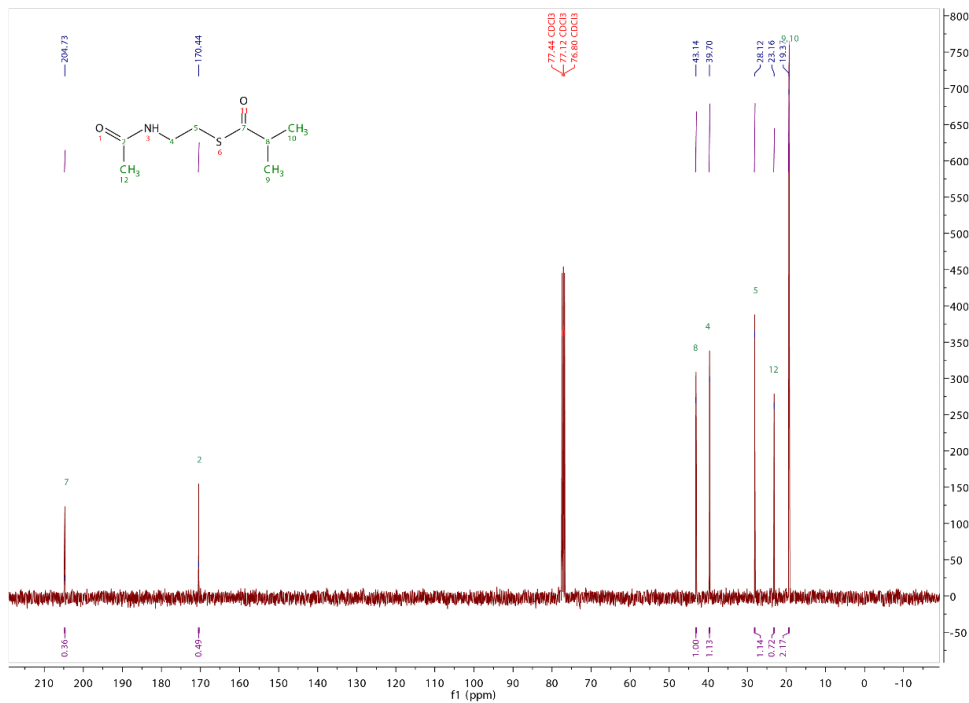
^1H NMR (400 MHz, Chloroform-*d*) δ 6.23 (s, 1H), 3.42 (q, $J = 6.3$ Hz, 2H), 3.01 (t, $J = 6.5$ Hz, 2H), 2.76 (hept, $J = 6.9$ Hz, 1H), 1.97 (s, 3H), 1.20 (d, $J = 7.0$ Hz, 6H).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 204.73, 170.44, 43.14, 39.70, 28.12, 23.16, 19.37.





Supporting Figure 3-10. ^1H NMR spectrum of isobutyryl-N-acetylcysteamine. CDCl_3 , 400 MHz.



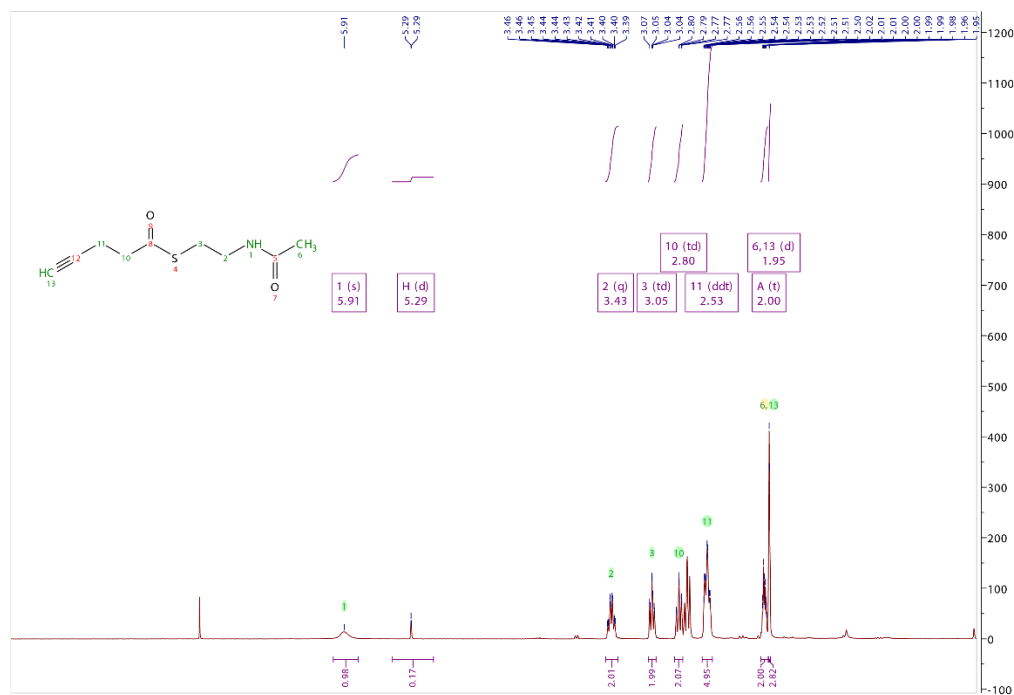
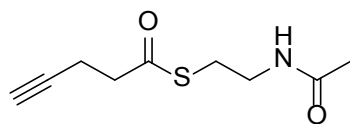
Supporting Figure 3-9. ^{13}C NMR spectrum of isobutyryl-N-acetylcysteamine. CDCl_3 , 101MHz.

S-(4-pentynoyl)-S-N-acetylcysteamine

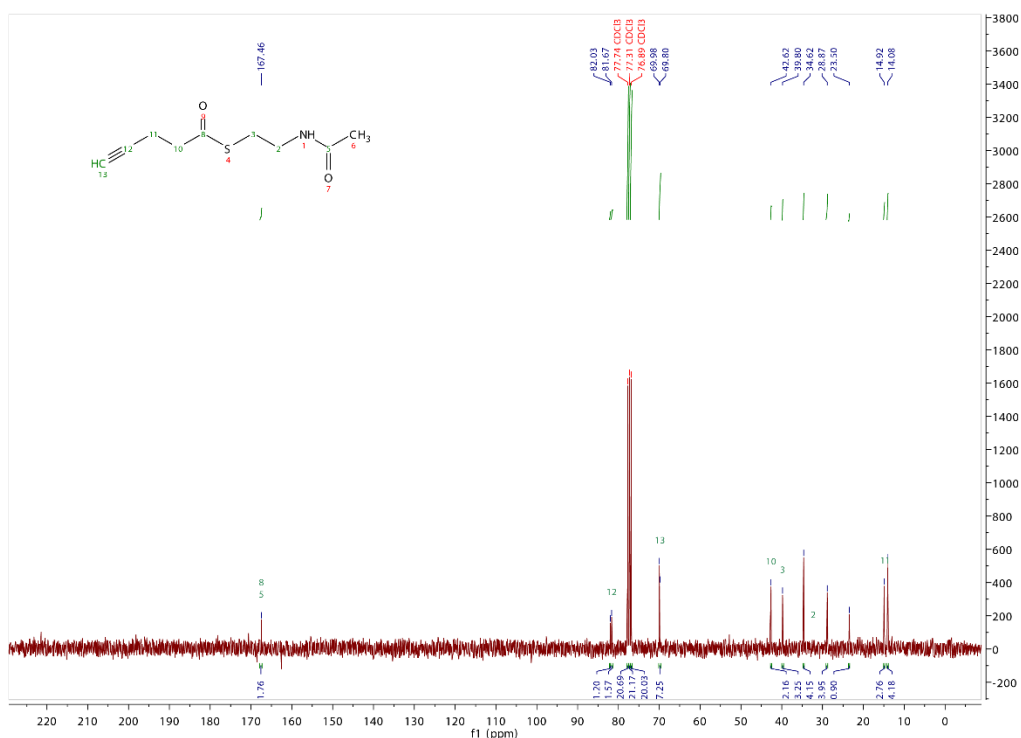
Into 2 mL dry tetrahydrofuran on ice was added, in order, 106 μ L N-acetylcysteamine (1 mmol), 288 mg EDC HCl (1.5 mmol), and 108 mg 4-pentynoic acid (1.1 mmol). The solution was stirred for 16 h, during which it warmed to room temperature. The mixture was diluted with 2 mL ethyl acetate and washed twice with 0.1 M aqueous HCl, twice with 5% aqueous sodium bicarbonate, and twice with brine. The organic phase was dried over sodium sulfate and evaporated under reduced pressure to afford a colorless oil. Rf = 0.5 on silica gel in ethyl acetate.

^1H NMR (300 MHz, Chloroform-*d*) δ 5.91 (bs, 1H), 3.43 (q, J = 6.2 Hz, 2H), 3.05 (td, J = 6.6, 1.6 Hz, 2H), 2.80 (td, J = 7.2, 1.5 Hz, 2H), 2.53 (dt, J = 7.2, 1.6 Hz, 2H), 1.99 (t, J = 2.6 Hz, 1H), 1.95 (d, J = 1.3 Hz, 3H).

^{13}C NMR (75 MHz, Chloroform-*d*) δ 198.16, 167.46, 82.03, 69.98, 42.62, 39.80, 28.87, 23.50, 14.92.



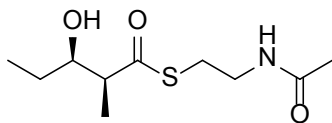
Supporting Figure 3-11. ^1H NMR spectrum of pentynyl-N-acetylcysteamine. Contaminants are due to residual 4-pentynoic acid. CDCl_3 , 300 mHz.



Supporting Figure 3-12. ^{13}C NMR spectrum of pentynyl-N-acetylcysteine. Contaminants are due to residual 4-pentynoic acid. CDCl_3 , 75 MHz.

S-(2S,3R)-3-hydroxy-2-methylpentanoyl-N-acetylcysteine

Provided by S. Yuzawa. Synthesized as reported in Yuzawa et. al 2017.(42)



3.4.5. NMR spectra

NMR spectra were acquired at the UC Berkeley College of Chemistry NMR facility on either Bruker Avance spectrometers, under partial funding from NSF grants CHE-0130862 and SRR023679A. Spectra were analyzed with MestReNova v11.0.4-18998.

3.4.6. Protein purification

BL-21(DE3) cells containing expression plasmids were pre-cultured overnight in 50 mL Terrific Broth (TB), $50 \mu\text{g mL}^{-1}$ kanamycin before being inoculated 1% v/v into 2 L TB media. Cells were at 37°C , 200 rpm until they reached $\text{OD}_{600} \geq .2$ AU and were then

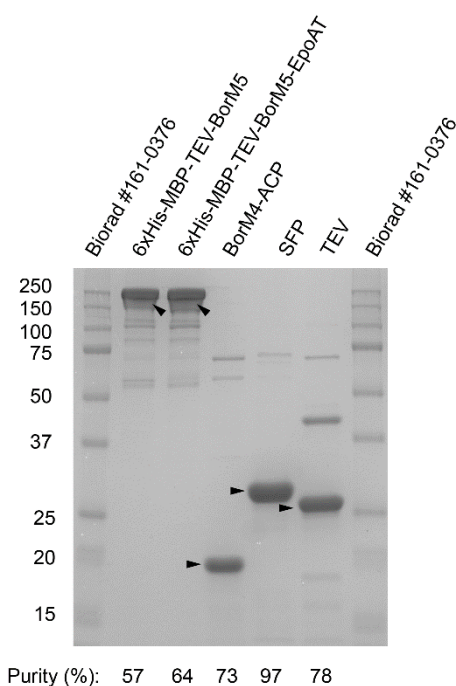
induced with 50 ng mL⁻¹ anhydrotetracycline or 250 mM isopropyl- β -D-galactopyranoside (IPTG) (Supporting Table 3-1). Cells were incubated at 18°C, 200 rpm for 20 hours. Cells were pelleted, the supernatant was discarded, and the cells were suspended in 50 mL lysis buffer (50 mM sodium phosphate, pH 7.6, 300 mM NaCl, 20 mM imidazole). Resuspended cells were lysed by 5-10 passes through an Avestin Emulsiflex C3 homogenizer (15,000 bar) and centrifuged (40,000 \times g, 30 min, 4°C) to pellet the insoluble debris. Co-NTA beads (Thermo) were equilibrated and used according to manufacturer specifications. Bound protein was eluted in with 20 mL elution buffer (50 mM sodium phosphate, pH 7.6, 200 mM imidazole). Co-NTA eluent was purified further with an Akta Explorer FPLC and a HiTrapQ anion exchange column according to manufacturer specifications. The target proteins were eluted by NaCl gradient at 1 mL*min⁻¹ between buffers A (50 mM sodium phosphate, pH 7.6 at 25oC, 8% glycerol, 1mM DTT) and B (50 mM sodium phosphate, pH 7.6 at 25oC, 500 mM NaCl, 8% glycerol, 1mM DTT). The correct fractions were identified by SDS-PAGE, pooled, concentrated and buffer exchanged into buffer A via spin filtration. Protein concentration was determined by Bradford assay (Bio-Rad, #500-0205). BorM5 was concentrated and buffer exchanged in TEV buffer (12.5 mM sodium phosphate, pH 7.6, 75 mM NaCl, 2.5 mM imidazole, 10 mM DTT) to a final volume of < 1 mL. TEV was added at a 1:50 molar ratio to 6xHis-MBP-TEV-BorM5 and incubated for 16 hours at 4°C. The reaction was diluted 1:10 in TEV buffer and re-purified by IMAC. Purified proteins were buffer-exchanged via spin filtration into Storage Buffer (50 mM sodium phosphate, pH 7.6, 1 mM DTT, glycerol 8% v/v), aliquoted (20 μ L), flash frozen with liquid nitrogen, and stored at -80°C.

Supporting Table 3-2. Proteins purified in this study.

Protein Name	Host	Induction	Purification	Purity (%)
6xHis-MBP-TEV-BorM5	BL-21 (DE3)	Anhydrotetracycline (50ng*mL ⁻¹)	NiNTA/AEC	57
6xHis-MBP-TEV-BorM5-epoAT	BL-21 (DE3)	Anhydrotetracycline (50ng*mL ⁻¹)	NiNTA/AEC	64
BorA4-ACP	BL-21 (DE3)	IPTG (250 μ M)	NiNTA/AEC	89
SFP*	BL-21 (DE3)	IPTG (500 μ M)	NiNTA	97

TEV protease	BL-21 (DE3)- CodonPlus RIL	IPTG (1mM)	NiNTA	78
---------------------	-------------------------------------	------------	-------	----

*Purified SFP was provided by Jacquelyn Blake-Hedges.



Supporting Figure 3-13. Purified proteins used. Arrows point to correct band.

3.4.7. SDS-PAGE of purified proteins

3ug purified protein was loaded onto a 12-well 8-16% polyacrylamide gel (Biorad #4561105) and run at 200V in TGS buffer (25 mM Tris pH 8.6, 192mM glycine, 0.1% SDS) until the loading dye exited the gel. Gels were Coomassie stained (Thermo 24594) (Supporting Figure 3-13). Protein purity was determined by densitometry using ImageJ. (216)

3.4.8. LCMS sample run and data acquisition

Samples were prepared for LC-MS/MS by chloroform methanol extraction, as reported. (217) Briefly, 50 μ L chloroform and 150 μ L water were added to methanol-quenched samples, and samples were vortexed between additions. Samples were centrifuged at 21,000 x g for 1 minute. The upper methanol-water was removed and 150 μ L methanol

was added. Samples were vortexed and centrifuged for 2 minutes at 21,000 × g. The chloroform-methanol layer was removed, and the residual liquid was evaporated in a fume hood for 1 hour. Re-solubilization of PKS proteins proved troublesome, so samples were incubated with 5 μ L DMSO for 30 minutes to aid solubilization. 45 μ L 100 mM ammonium bicarbonate, 20% methanol was added. Samples were trypsinized at 1:100 w/w porcine trypsin for 16 hours at 37°C. Samples were centrifuged for 5 minutes at 21,000 × g to pellet insoluble protein. The supernatant was removed, and protein concentration was quantified by 280 nm absorption on a Nanodrop (Thermo). Peptide samples were analyzed using an Agilent 1290 Infinity II liquid chromatography system coupled to an Agilent 6460 QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA). The peptide samples (5 μ g) were separated on an Ascentis Express Peptide ES-C18 column (2.7 μ m particle size, 160 Å pore size, 50 mm length × 2.1 mm i.d.), couple to a guard column (5 mm × 2.1 mm i.d.) with similar particle and pore size (Sigma-Aldrich, St. Louis, MO). The system was operated at a normal flow rate of 400 μ L*min⁻¹ and column compartment heated at 60 °C. Peptides were eluted into the mass spectrometer via a gradient with initial starting condition of 95% Buffer A (0.1% formic acid) and 5% Buffer B (99.9% acetonitrile, 0.1% formic acid). Several gradients were used throughout the course of these studies. These gradients, and the experiments to which they correspond are enumerated (Supporting Table 3-3 and Supporting Table 3-4, respectively). The peptides were ionized by an Agilent Jet Stream ESI source operating in positive-ion mode with the following source parameters: Gas Temperature = 250 °C, Gas Flow = 13 L*min⁻¹, Nebulizer Pressure = 35 psi, Sheath Gas Temperature = 250 °C, Sheath Gas Flow = 11 L*min⁻¹, and VCap = 3,500 V. Collision energies were predicted by Skyline, except for the phosphopantetheine ejection parent-daughter ion pairs, which were all set to 30eV. Dwell time = 20 ms. The data were acquired using Agilent MassHunter Data Acquisition, version B.08.00 (Build 8.0.8023.5 SP1).

Supporting Table 3-3. LC gradients

Gradient 1			Gradient 2		
Time (min)	Buffer A	Buffer B	Time (min)	Buffer A	Buffer B
0	95	5	0	95	5
0.2	95	5	4	95	5
5.7	65	35	9.7	65	35
6	10	90	10	10	90
8	10	90	12	10	90
8.5	95	5	12.5	95	5
13	95	5	17	95	5

Supporting Table 3-4. LC gradients used for each experiment.

Figure	Datum	Gradient	Figure	Datum	Gradient
3a	Isobutyryl-ACP4 trace	1	5a	Pentynyl-SNAC	2
3b	Isobutyryl-ACP4 time course	1	5a	Diketide-SNAC	2
3c	Isobutyryl-ACP4 concentrations	1	5a	Hexanoyl-TP	2
4a	Isobutyryl-ACP4	2	Raw data	Dodecanedioic-TP	2
4a	Butyryl-ACP4	2	6a	BorM5 Epo-AT	1
4a	Hexanoyl-ACP4	2	S2	Preparation of acyl-ACP4	1
4a	3-hydroxybutyryl-ACP4	2	S3	Activity of TEV-cleaved BorM5	2
Raw data	Decanoyl-ACP4	2	S4	Targeted LC-MS/MS of partially reduced intermediates	2
Raw data	Myristoyl-ACP4	2	S5	Semi-quantitative yields	2
5a	Acetyl-SNAC	2	S6	Isobutyryl-SNAC time course	1
5a	Propionyl-SNAC	2	S7	Latent addition of substrates	1
5a	Isobutyryl-SNAC	2	S8	Re-addition of substrates	1

3.4.9. LC-MS/MS Phosphopantetheine ejection assay

Skyline (218) was used to generate targeted tandem mass spectrometry methods for the phosphopantetheine ejection assay. (202) The Skyline method, including the full list of

parent-daughter ion pairs is available on Panorama. As Skyline is not natively set up for the phosphopantetheine ejection assay, we developed a work-around. In brief;

- ACP tryptic fragments containing phosphopantetheine attachment sites (BorM5: DLGFD~~S~~LTAVELR, active serine in red) were modified to include the acyl-phosphopantetheine mass (H₂₁C₁₁N₂O₆PSH: 341.09Da) on the terminal arginine or lysine.
- A neutral loss was added corresponding to the terminal arginine or lysine, covalently bound to phosphate (C₆H₁₅N₄O₂PO₃: 254.078Da in the case of arginine).
- Daughter ions were chosen as y1 - (neutral loss).

See Raw Data files for examples.

3.4.10. Raw data

Raw data for all main text figures are available on the targeted MS/MS data sharing platform Panorama. (219) Additionally, the traces where no intermediates were identified are available.

<https://panoramaweb.org/project/JBEI/SC%202018%20Probing%20the%20flexibility%20of%20an%20iterative%20modular%20polyketide%20synthase%20with%20non-native%20substrates%20in%20vitro/begin.view?>

3.4.11. Data analysis

Data files were analyzed using Skyline version 3.6.0.10493 (MacCoss Lab, University of Washington, Seattle, WA).(218) Substrate peaks were identified by their absence from the no-substrate controls as well as their characteristic distribution, i.e. intermediate length correlated positively with retention time (Figure 3-3, panel a). Peaks were integrated and normalized to control peptides that are present in BorM5, but do not participate in the reaction. Normalized peak areas were background subtracted from no-substrate control at the same retention time (Supporting Equation 3-1). Normalized peak areas for observable extension intermediates were then summed and represented as a percentage of the sum (Supporting Equation 3-2). Peaks were assumed to be comparable (i.e. no changes in ionization efficiency) as they are fully reduced extensions of the same starting substrate.

Supporting Equation 3-1. Operations for normalization of peak area to control peptides and background subtraction.

$$\text{Normalized peak area (A.U.)} = \text{NPA(A.U.)} = \frac{\text{Peak area}_{\text{acyl-ACP, treatment}}}{\text{Peak area}_{\text{control peptide, treatment}}} - \frac{\text{Peak area}_{\text{acyl-ACP, no substrate control}}}{\text{Peak area}_{\text{control peptide, no substrate control}}}$$

Supporting Equation 3-2. Operations for normalization of peaks areas as percentage of the sum extension intermediates.

$$\text{NPA (\%)}_{\text{extension } i} = \frac{\text{NPA(A.U.)}_{\text{extension } i}}{\text{NPA(A.U.)}_{\text{extension } i} + \text{NPA(A.U.)}_{\text{extension } i+1} + \dots + \text{NPA(A.U.)}_{\text{extension } n}}$$

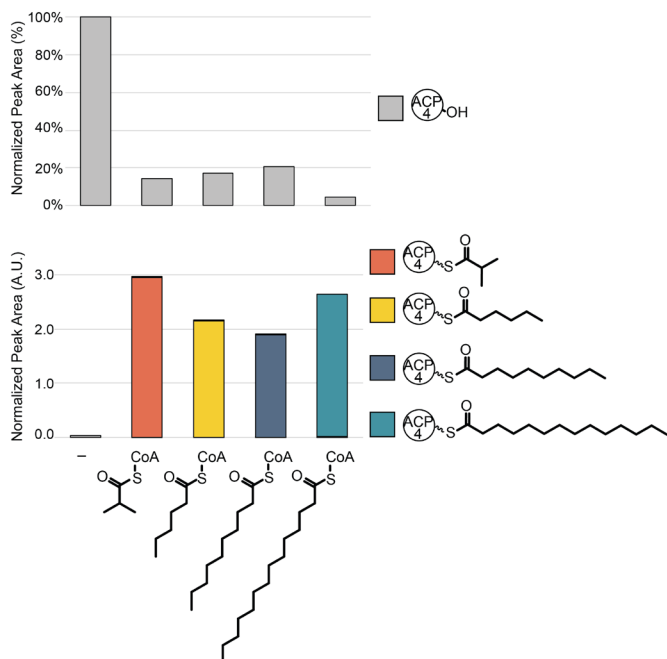
3.4.12. Phosphopantetheinylation assays (acyl-ACP formation)

Unless otherwise stated, 40 μL 10 μM *apo*-ACP4 or *apo*-6xHis-MBP-BorM5 was incubated with 250 μM acyl-CoA, 1 μM *Bacillus subtilis* Sfp, 10 mM MgCl, 2.5 mM TCEP, and 100 mM sodium phosphate, pH 7.6 overnight at 25°C.

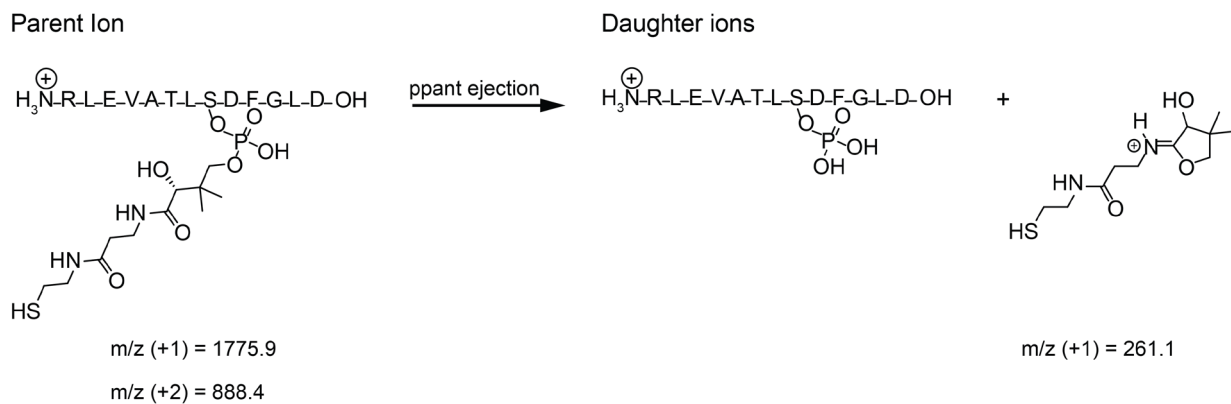
3.4.13. Extension assays

Unless otherwise stated, 40 μL 10 μM *holo*-BorM5 was incubated with either (1) 10 mM acyl-SNAC, or (2) 10 μM acyl-ACP4, and 500 μM methylmalonyl-CoA and 1 mM NADPH. Final reaction volumes were 80 μL . In the case of the extension assays with malonyl- and methylmalonyl-CoA, both substrates were provided at 500 μM . Reactions were quenched by the addition of 200 μL LC-MS grade methanol.

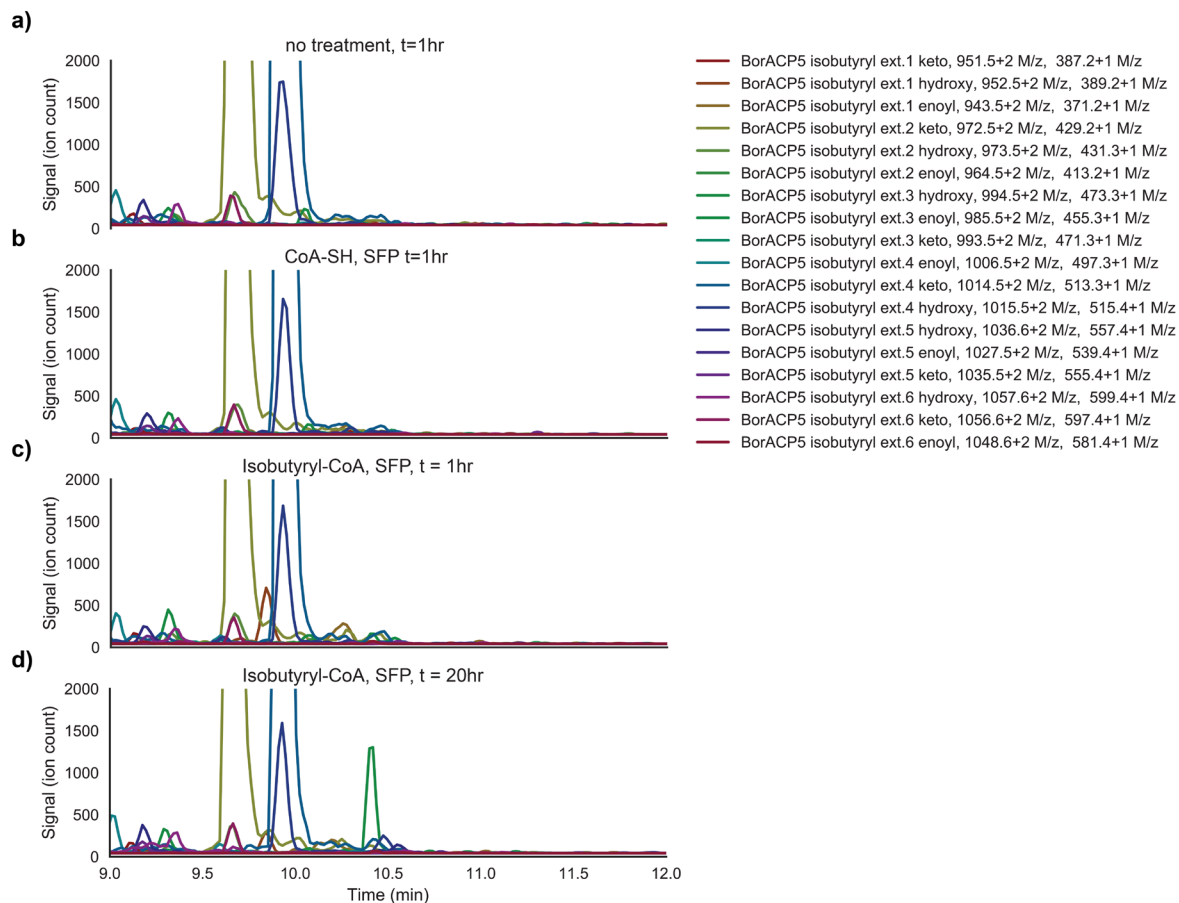
3.5. Supporting Figures



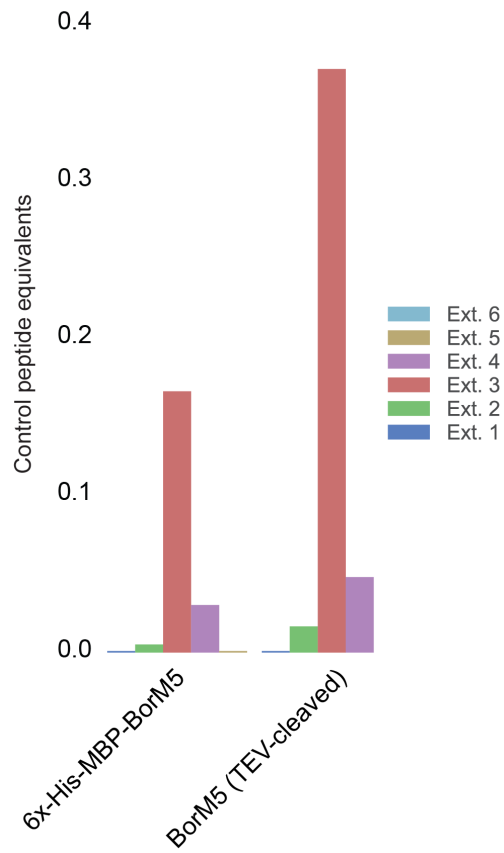
Supporting Figure 3-14. Overnight phosphopantetheinylation of apo-ACP4. Apo-ACP4 was phosphopantetheinylated as stated in the main text. a) Relative peak areas for acyl-ACP4 identified by LC-MS/MS (top). b) Relative quantification of the acyl-ACPs is deduced by the loss of the apo-ACP4 compared to the no-substrate control.



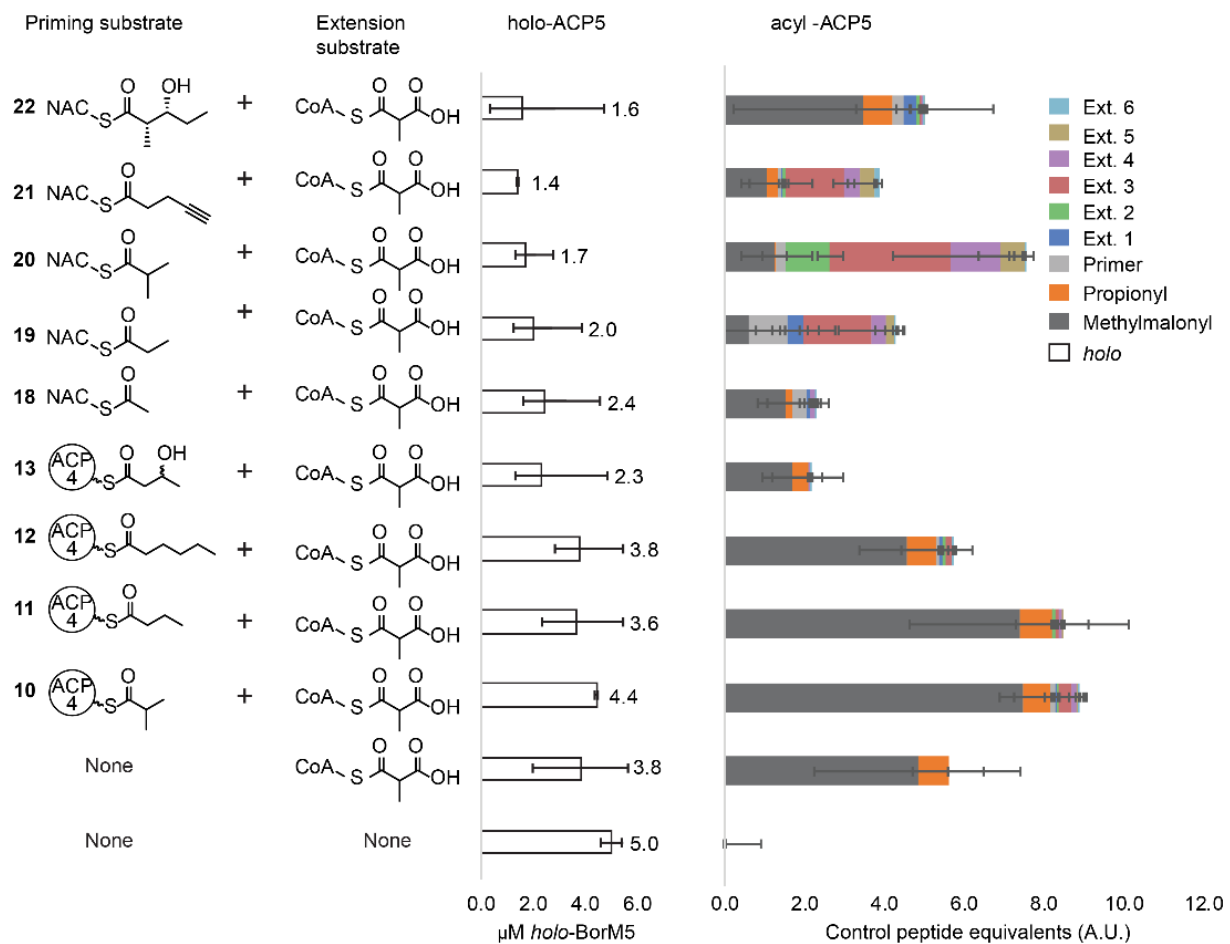
Supporting Figure 3-15. Phosphopantetheine ejection assay. The 'parent' ion is the tryptic ACP peptide containing the DSL active site motif, the phosphopantetheine prosthetic arm, and the acyl group, if present. The 'daughter' ion is phosphopantetheine fragment and acyl group



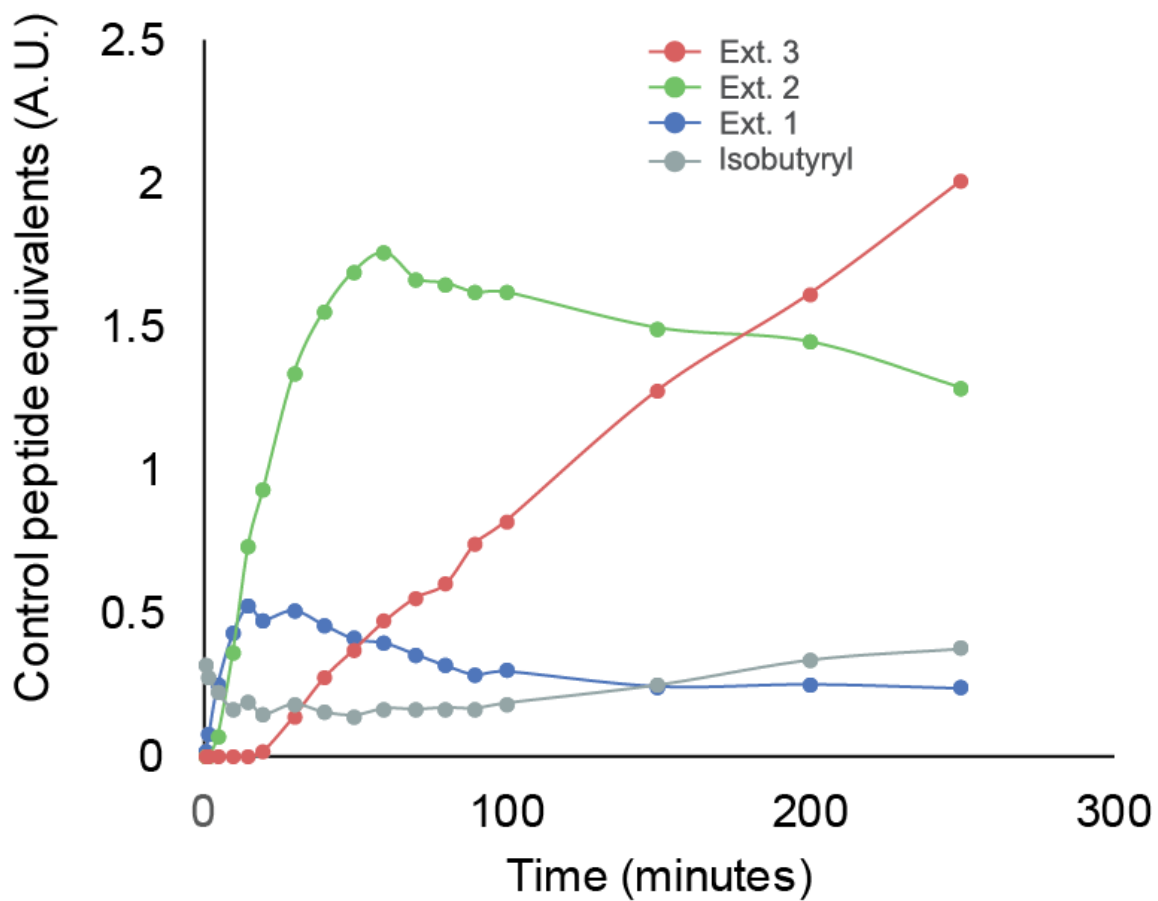
Supporting Figure 3-16. LC-MS/MS traces targeting partially reduced intermediates bound to BorM5-ACP with either **a)** apo-ACP4 (no treatment), **b)** holo-ACP4, **c)** isobutyryl-ACP4 with methylmalonyl-CoA, and NADPH, t=1hr, and **d)** isobutyryl-ACP4 with methylmalonyl-CoA, and NADPH, t=20hrs.



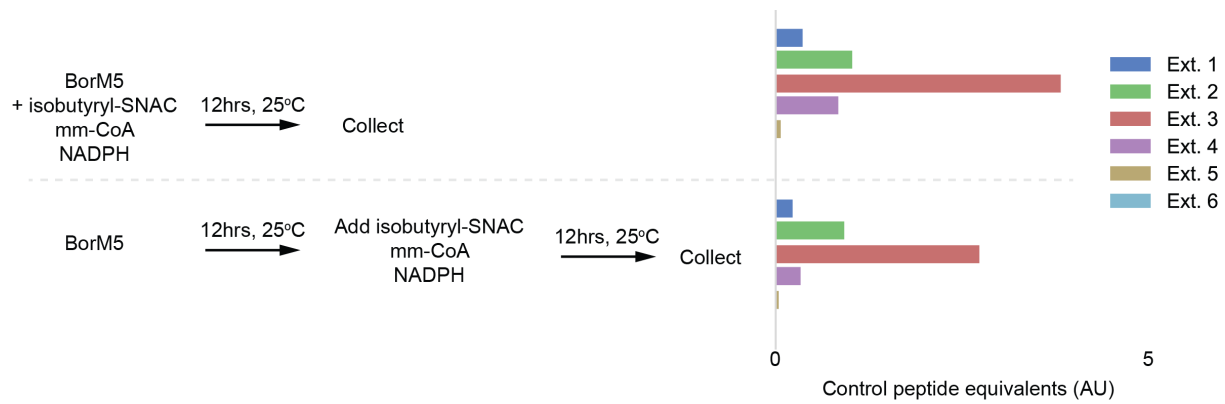
Supporting Figure 3-17. LC-MS/MS intermediate analysis of tagged and cleaved BorM5 with isobutyryl-ACP4, methylmalonyl-CoA, and NADPH after 8 hours.



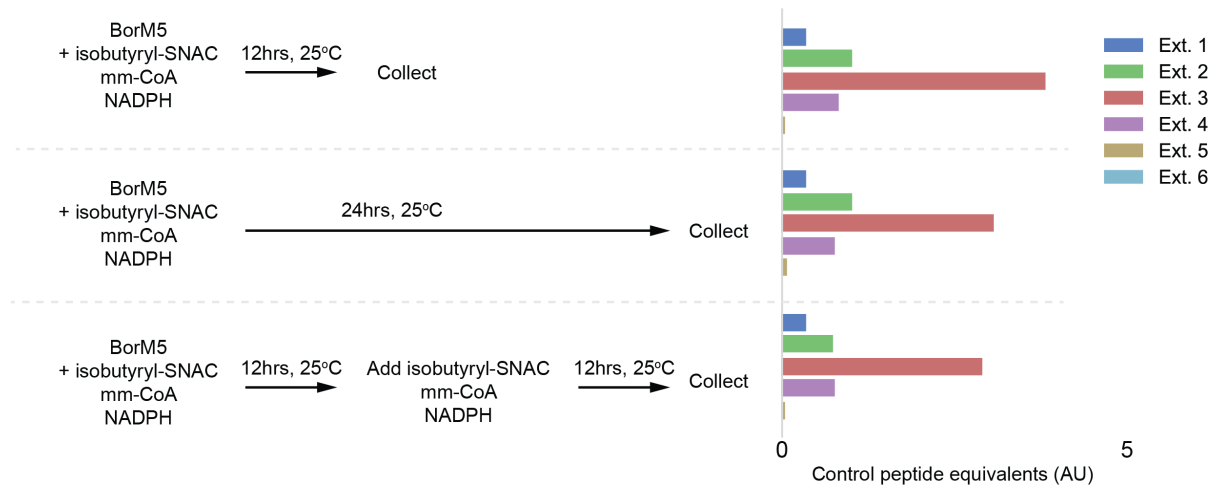
Supporting Figure 3-18. Semi-quantitative yields of acyl-BorM5 by loss of holo-BorM5. Error bars are standard deviation between technical triplicate.



Supporting Figure 3-19. LC-MS/MS time course assay of BorACP5-bound intermediates with 10 mM isobutyryl-SNAC, methylmalonyl-CoA, and NADPH. Time points are 1, 2, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, and 250 minutes.



Supporting Figure 3-20. LC-MS/MS intermediate profiles of BorM5 and isobutyryl-SNAC at t=12hrs (top) or after 12 hours of incubation at 25C and 12hours with substrates (bottom).



Supporting Figure 3-21. LC-MS/MS intermediate profiles of BorM5 and isobutyryl-SNAC at t=12hrs (top), t=24hrs (middle), and at 24 hours with re-addition of substrates after 12 hours (bottom).

3.6. Acknowledgements

The authors thanks J. Blake-Hedges for providing purified SFP. pRK793 was a gift from D. Waugh.

3.7. Funding Sources

This work was funded by the National Science Foundation, *via* Awards CBET-1437775, MCB-1341894 and EEC-0540879 to the Synthetic Biology Engineering Research Center, and was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. This research was also conducted as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies and Vehicle Technologies Offices. The NMR work presented here was performed on instruments funded in part by NIH grant SRR023679A and NSF grant CHE-0130862.

Chapter 4. Structure, function, and engineering of BorB, the type II thioesterase from the borrelidin biosynthetic cluster

4.1. Abstract

α/β hydrolases are a large and diverse protein superfamily. In Natural Product biosynthesis, α/β hydrolases called thioesterases can terminate biosynthetic assembly lines by releasing the product. Thioesterases can also act in *trans* and are thought to remove aberrant intermediates and restart stalled biosynthesis. The borrelidin biosynthetic gene clusters from *Streptomyces parvulus* contains a hitherto uncharacterized standalone thioesterase, BorB. We sought to use BorB to release long-chain intermediates from BorM5, however biochemical assays demonstrated that BorB has a preference for propionate. A crystal structure of BorB shows a wedge-like a substrate-binding crevice that limits substrate length. To investigate the structure-function relationship, we made chimeric BorB variants using loop regions from characterized homologs with varying specificity. These chimeras led to modest shift in activity. Combined these data support BorB's classification as an editing thioesterase. The structure-function relationships described here are essential for understanding and applying *trans* acting thioesterases.

4.2. Introduction

Having demonstrated the ability of BorM5 to produce branched-chain acyl-ACP, we sought an enzyme to offload intermediates as free fatty acids. In addition to the *cis*-TE, many BGCs also contain a second, *trans*-acting type II TE (TEII) that operates distinctly from the biosynthetic assembly line. Knockout of TEIIs in vivo often leads to reduced product yields. (207) TEIIs are recognized to generally act as editing TEs that remove aberrant intermediates from ACPs. (207) "Dead-end" intermediates arise through several mechanisms including the decarboxylation and protonation of malonyl-CoA and its analogs or the erroneous acyl-phosphopantetheinylation of ACPs. (220) As such, the majority of the studied TEIIs are specific for short-chain acyl substrates, such as acetyl- and propionyl-ACP. (207) However, some TEIIs play more specialized roles in starter unit control, provision of intermediates, and product release. (207) Well-characterized examples include the TEIIs from the tylosin, (221), FR-008, (222) coelimycin, (223) pikromycin, (224), nanchangmycin, (225), tyrocidine, (226), surfactin, (227, 228), rifamycin, (229), prodigiosin, (230) surfactin, (227, 228), and colibactin BGCs. (231)

The borrelidin cluster contains a putative TEII, BorB. (175) Genomic disruption of BorB reduced the yield of borrelidin by 25%. (175) To investigate BorB's biochemical

activity, we characterized the purified enzyme in vitro with acyl-SNACs, acyl-CoAs, and acyl-ACPs. We discover that BorB readily hydrolyzes short-chain aliphatic substrates but has limited acceptance of longer fatty acids and diacids. Additionally, we derived an X-ray crystal structure of BorB which reveals a canonical α/β hydrolase fold with a clear substrate binding cleft. The amino acids lining this cleft are similar to other TEIIs with editing activity. To investigate the structure-function relationship in BorB, we made multiple loop-swapped chimeric proteins. Most mutants resulted in loss of activity, but several exhibited modest improvements for C2 and C3 substrates. Overall, these findings confirm BorB's identity as a TEII and support its role as an editing enzyme.

4.3. Results

4.3.1. BorB is a TEII

BorB clades with the TE18 family of α/β hydrolases from PKSs and NRPSs based on primary sequence alignment. (232) Alignment of over 200 putative and well-studied TEII such as RifR and RedJ, revealed a classic catalytic triad of S98, D204, and H232 (numbering based on our cleaved expression construct) (See Methods). To confirm BorB's thioesterase activity, we expressed codon optimized BorB in *E. coli*. We could only detect soluble BorB when expressed as a hexahistidine maltose binding protein (MBP) fusion with a tobacco etch virus (TEV) protease site preceding BorB. (200) 6xHis-MBP-TEV-BorB was purified, cleaved from its tag, and reacted with synthesized acyl-N-acetylcysteamine thioesters (SNACs), well-established substrate analogs for PKSs and TEIIs. (233) We continuously monitored thioester cleavage via the colorimetric Ellman's assays (Supporting Figure 4-7). (234) BorB readily cleaved acetyl-, and hexanoyl-SNAC and the activity was linearly dependent on both time and enzyme concentration (Supporting Figure 4-8, B). To verify that specifically BorB, and not a co-eluent, was responsible for the observed cleavage, we mutated the catalytic serine to an alanine. This mutant, BorB S98A, had no detectable activity when purified and reacted with acetyl-SNAC (Supporting Figure 4-9).

4.3.2. BorB substrate specificity

To investigate BorB's substrate preference, we performed Michaelis-Menten kinetic assays with aliphatic acyl-CoA substrates (Supporting Equation 4-1). As we were limited by substrate solubility, we used the linear region of rate curves to estimate K_{cat}/K_m (Supporting Equation 4-2). Hydrolysis rates are shown in Figure 4-1 the rates are tabulated in Table 4-1.

BorB had the highest activity on propionyl-CoA ($K_{cat}/K_m = 22.4 \pm 3.5 \text{ M}^{-1} \text{ min}^{-1}$), followed by hexanoyl-CoA ($K_{cat}/K_m = 8.9 \pm 3.1 \text{ M}^{-1} \text{ min}^{-1}$). After hexanoyl-CoA, activity correlated negatively with chain length; octanoyl- and decanoyl-CoA were cleaved with K_{cat}/K_m of 4.5 ± 2.5 , and $1.2 \pm 1.0 \text{ M}^{-1} \text{ min}^{-1}$, respectively. BorB had low activity ($1.7 \pm 1.3 \text{ M}^{-1} \text{ min}^{-1}$) on acetyl-CoA. We did not detect thioester cleavage above background for palmitoyl-CoA, isobutyryl-CoA, 3-hydroxybutyryl-CoA, and diacids malonyl-, methylmalonyl-, succinyl-, or adipyl-CoA at 2 mM concentrations (Supporting Figure 4-10).

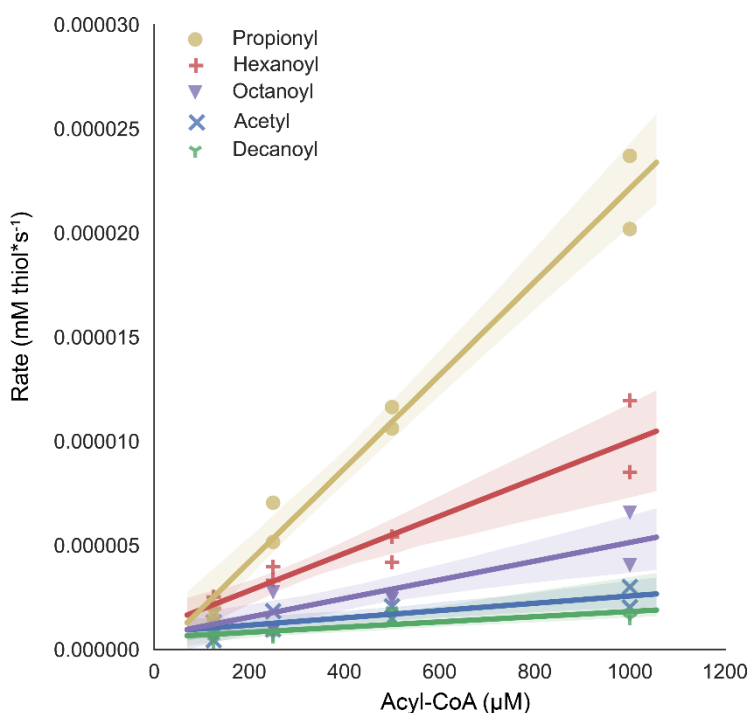


Figure 4-1. Kinetic assays for BorB-mediated hydrolysis of A) acyl-SNACs and B) acyl-CoAs.

Table 4-1. Kinetic parameters for BorB. N.d. = not detected. N = 2 experiments of technical triplicate on different days.

Acyl-CoA	K_{cat}/K_m ($M^{-1} s^{-1}$)
Acetyl	1.7 ± 1.3
Decanoyl	1.2 ± 1.0
Hexanoyl	8.9 ± 3.1
Octanoyl	4.5 ± 2.5
Propionyl	22.4 ± 3.5
Palmitoyl	N.D.
Malonyl	N.D.
Methylmalonyl	N.D.
Succinyl	N.D.
Adipyl	N.D.
3-hydroxybutyryl	N.D.
Isobutyryl	N.D.

As TEIs generally interact with ACPs and PCPs of biosynthetic assembly lines, the probable native substrate of BorB is an acyl-ACP. (207) To test if BorB cleaved acyl-ACPs, we incubated BorB with purified mono-domain acyl-BorM5-ACP, from the fifth module of the borrelidin synthase. Acyl-ACPs were identified by matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS). Despite the rather poor mass resolution (± 10 Da of the expected value), we observed clear peak shifts from holo-acyl-ACP to acyl-ACPs ranging from C2-C14 (Figure 4-2). After one hour of incubation with BorB, we observed cleavage of acyl-BorM5-ACP with chains up to decanoyl (C10). We did not observe any cleavage of myristoyl-BorM5-ACP (C14).

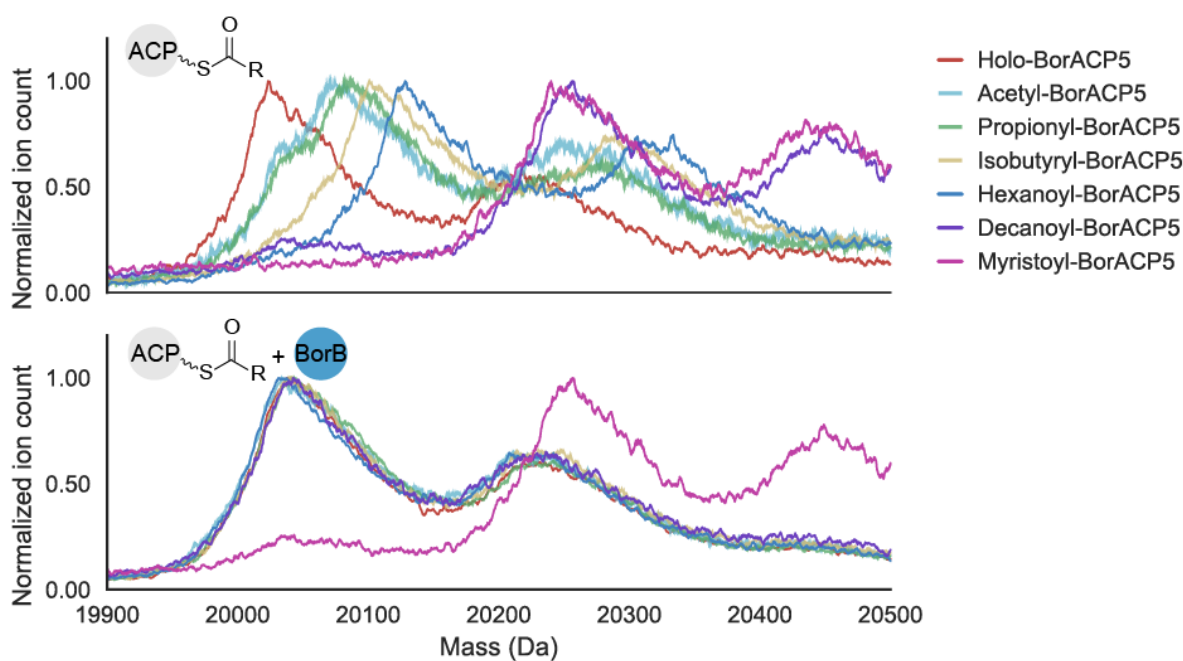


Figure 4-2. MALDI-TOF of acyl-BorACP5 without (top) and with (bottom) 1 μ M BorB after one hour. Colors correspond to different acyl chains. T = 1 hour.

4.3.3. Structure of BorB - overall structure

An X-ray crystal structure of BorB was solved to 1.93 Å (See section Methods, Supporting Table 4-4). BorB exhibits the predicted α/β hydrolase fold. (235) Two molecules of BorB formed the asymmetric unit (Supporting Figure 4-13). However, BorB lacks the N-terminal helices reported to be important for dimerization in the Pik and DEBS TEs (Supporting Figure 4-11). (236) To confirm BorB's quaternary state, we performed size-exclusion chromatography coupled to small angle X-ray scattering (SEC-SAXs). The scattering curve defines a monomeric globular protein with a radius of gyration (R_G) of 19 Å and a D_{max} of 64 Å (Supporting Figure 4-3). Back-calculation of scattering curves was calculated using the FOXS server. (237, 238) The experimental data are more consistent with monomeric BorB than dimeric BorB ($\chi^2_{monomer} = 1.08$, $\chi^2_{dimer} = 19.85$) (Supporting Figure 4-12). Thus, purified BorB is monomeric.

BorB forms a two-lobed structure with an α/β hydrolase core and a "lid region" (Figure 4-3, A). The core begins with a pair of antiparallel β -strands ($\beta 2$ - $\beta 4$) and continues with alternating α -helices and β -strands. BorB has seven structured β -strands (2- 8) and lacks the $\beta 1$ strand, similar to other PKS TEs (Figure 4-3, C). (153) The lid region, while found commonly in α/β hydrolases, contributes significantly to structural variability among these synthases. In BorB and other monomeric TEs, the lid extends from between strands $\beta 6$ and $\beta 7$ (Figure 4-3, C). (153) BorB's lid is formed from three α -helices ($\alpha L1$ -3, res. 138-183). The catalytic triad - Ser98, Asp204, and His232 - sits at the base of the crevice formed between the lid and the core subdomains. The dimers differ slightly in several regions; the electron density for the A and B chains begin at T8 and N7, respectively. The C-termini differ by one residue as well. The largest difference is in the lid region; Chain B contains an intact lid, while Chain A lacks electron density for the C-terminal half of $\alpha L2$ (Figure 4-3, B). These forms, dubbed "Open" and "Closed" conformations, have been captured in numerous TE structures. (86, 229–231, 239) The lid and its associated flexible loops likely control the accessibility substrate binding chamber to acyl-ACPs via an N-terminal ACP entrance. In BorB's Open configuration, the catalytic cleft is accessible from both sides. In the Closed configuration, the N-terminal ACP entrance is obstructed by $\alpha L2$ of the lid. Additionally, the lid shows considerably flexibility in both chains. The lid region also differs more between monomers than the core ($RMSD_{Core} = 1.178\text{\AA}$, $RMSD_{Lid} = 2.075\text{\AA}$). B-factors, indicating chain flexibility, are highest between helices $\alpha L2$ and $\alpha L3$ (Supporting Figure 4-14).

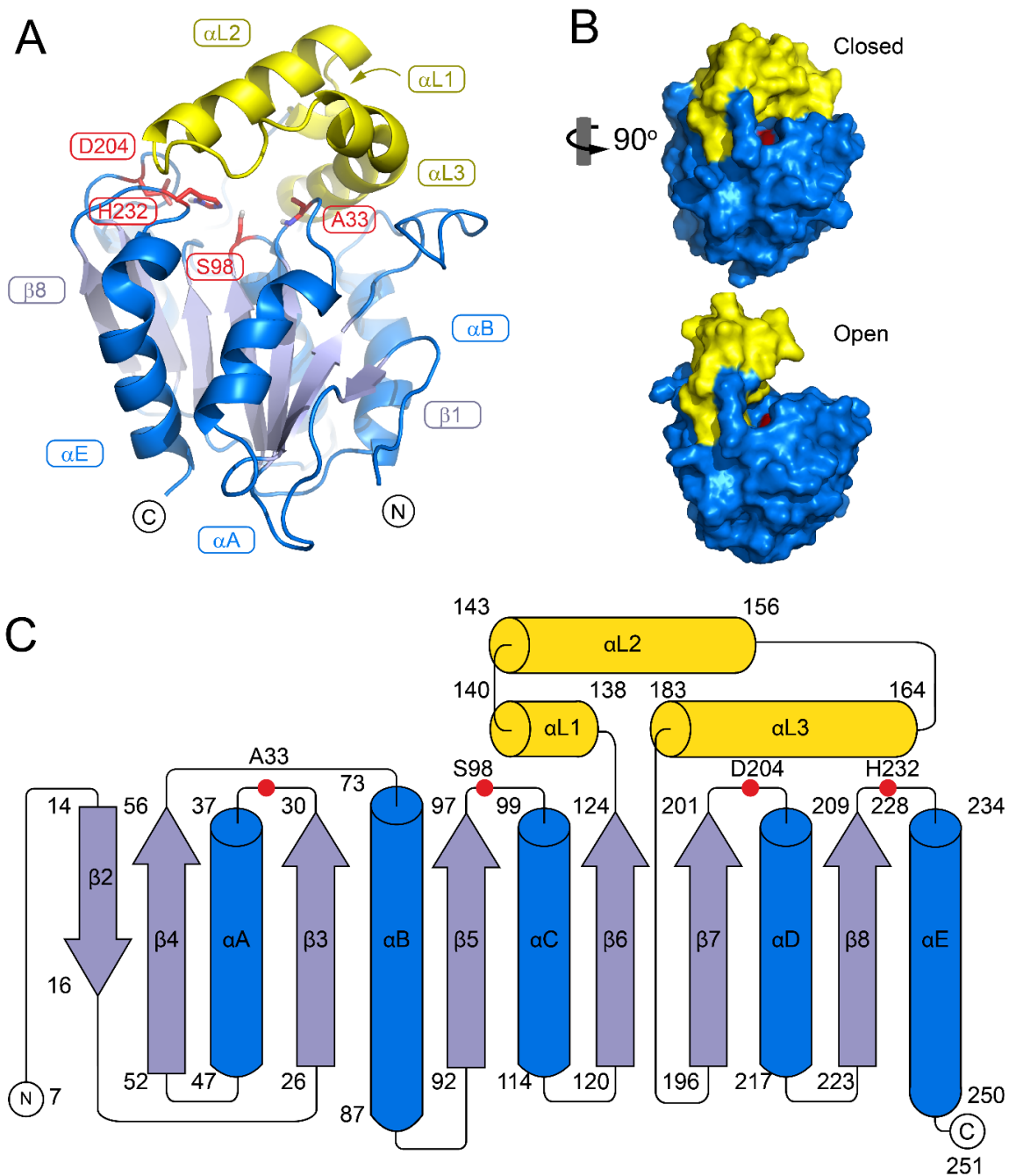


Figure 4-3. BorB overall structure. **A)** Overlaid structure of the A and B chains of BorB. Lid regions are colored orange and yellow, respectively. Active triad is colored in red. **B)** Individual monomers of the asymmetric unit. Coloration is identical to A. **C)** Topology diagram of BorB.

4.3.4. Structure of BorB - catalytic site

The lid and its associated flexible loops control the size, shape, and presumably substrate-binding preference of the chamber. In the macrocyclizing TEs from the pikromycin and DEBS synthases the substrate channel penetrates through the enzyme, forming a donut-like aperture that favors macrocyclization. (236, 240) In contrast, the substrate channel in BorB resembles a wedge-like crevice that is also found in many non-cyclizing TEs (Figure 4-3, A). (86, 229–231) The catalytic serine sits in the canonical G96-H97-S98-X99-G100 motif after strand β 5. This residue is positioned in a hairpin loop at N-terminus of helix α C. This “nucleophilic elbow” is typical of α/β hydrolases. (241) The other members of the catalytic triad, D204 and H232, sit after strands β 7 and β 8, respectively, and are hydrogen bonded. The backbone amide of A33 forms the putative oxyanion hole. (229)

4.3.5. Structure of BorB - substrate binding chamber

Numerous BorB homologs have been modeled or co-crystallized with substrates, non-hydrolyzable analogs, and ACPs (Supporting Figure 4-15). Studies of the DEBS TE identified an arginine-rich patch on the N-terminal half of the protein, to which the ACP was computationally docked. (236) Structures the surfactin TE with bound peptides, (239) hFAS TE with polyunsaturated fatty acids, (242) the pikromycin TE with non-hydrolysable substrate analogs, (243) co-crystallization of the NRPS enterocin thioesterase EntF with apo-(244–246) and crypto-ACP, (244–246) and co-crystallization of the valinomycin TE Vlm2 with peptide intermediates support this model, (246) and lead to a putative chamber for the acyl substrate where the acyl-ACP enters from the right side of the crevice (Supporting Figure 4-16). The substrate binding chamber therefore lies to the left of S98.

In BorB, the front of this putative binding chamber is bounded by the “Catalytic Loops” 1 and 2, which contain the catalytic D204 and H232, respectively (Figure 4-4, A). Catalytic Loop 2 is highly conserved (Supporting Figure 4-17), while Catalytic Loop 1

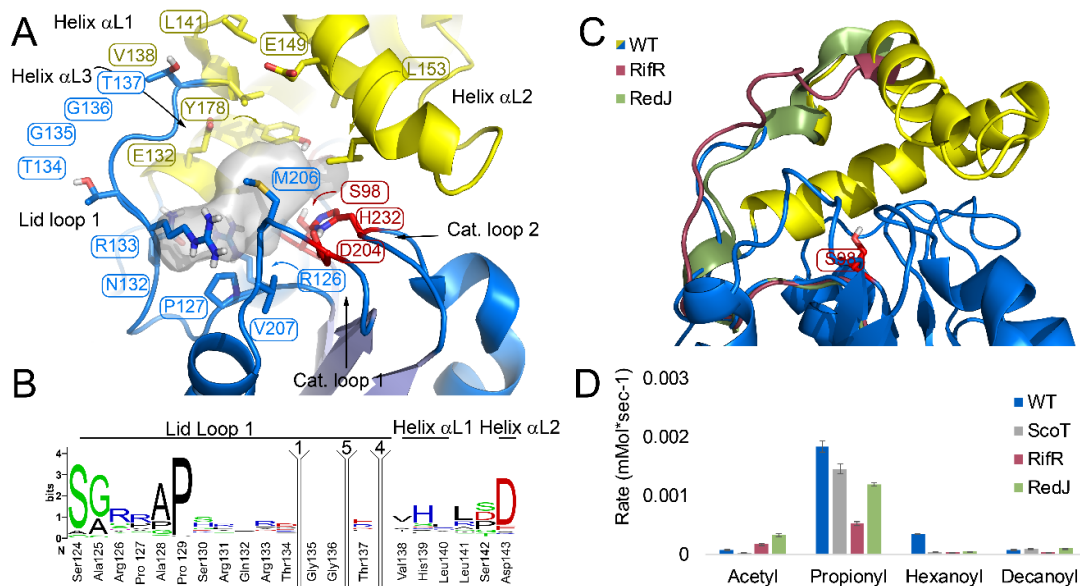


Figure 4-4. **A**) Active site of BorB. Cavity shown in grey surface. **B**) Conservation of the region used for generating chimeric proteins. Larger letters indicate higher conservation. Collapsed regions are indicated by the number of amino acids and are not present in BorB. Gly135 and Gly136 are unique to BorB. **C**) Overlay of BorB B chain with RifR (3FLA) and RedJ (3QMW) structures. The swapped loop regions are colored. **D**) Acyl-CoA turnover assay for relative WT activity. Error bars are standard deviation of technical duplicate.

contains considerable diversity apart from catalytic D204, (Supporting Figure 4-17). The binding pocket is largely hydrophobic proximal to S98, with contributions from residues M99, A33, and A181.

The pocket bifurcates towards the “Exit” of the tunnel. The top and back boundaries are formed by helix α L2 and α L3, respectively (Figure 4-4, A). In the long-chain hydrolyzing RedJ, this pocket is lined with short hydrophobic residues and one form of RedJ was co-crystallized with PEG in the crevice. (230) However, in the editing TEII RifR, this pocket is occluded by bulky, polar residues. (229) Similarly, the pocket in BorB is lined with polar residues, D182, Y178 and E149. The proposal that these residues differentiate editing TEs from long-chain hydrolyzing TEs is consistent with BorB’s activity. Lid Loop 1 and helix α L1 form the left wall of the pocket (Figure 4-4, A). This wall is made of 20-30 poorly-conserved residues bounded by a conserved serine and a conserved aspartate (Figure 4-4, B). While there is low conservation at these positions, the most common residues are also positively charged (Figure 4-4, B).

4.3.6. Engineering BorB

Lid loop 1 has been proposed to control substrate preference in several TEs. (86, 230, 231). To investigate this possibility, we generated chimeric swaps with characterized TEIIs RifR, RedJ and ScoT. (223, 229, 230) RifR and RedJ prefer C10-C12 esters while ScoT prefers propionate esters. RifR and RedJ have been structurally characterized. All four proteins contain Lid Loop 1 (Figure 4-4 C) and differ considerably in this region (Supporting Figure 4-18).

The chimeras were purified from *E. coli*. BorB-RifR and BorB-ScoT had roughly 25% yield of WT BorB, indicating instability and degradation. These proteins were assayed for altered activity on acyl-CoAs using Ellman's assay (Figure 4-4 D). Donor loops from ScoT and RedJ lead to ~30% loss in activity on propionyl-CoA while the loop from RifR lead to a 75% loss in activity. BorB-RifR and RedJ mutants exhibited a modest 2-4-fold increase in activity with acetyl-CoA. However, no mutants had improved activity on hexanoyl- or decanoyl-CoA.

4.4. Discussion

BorB's broad substrate preference, combined with the fact that it improves but is not required for borrelidin biosynthesis, suggests that it is an editing TEII. (175) There are two predominant models for TEII specificity: in a high specificity model, the TEII shows high activity for a particular substrate which is most commonly a short-chain byproduct of malonyl-ACP decarboxylation. In the low specificity model, the TEII does not prefer any particular substrate, but has a low level of promiscuous activity for substrates varying in chain length and substitution. (207) BorB fits into this second category. Of the acyl-CoAs, BorB exhibited the highest activity on propionyl-CoA. Propionyl-CoA can arise from the decarboxylation of methylmalonyl-CoA and aberrant protonation, rather than condensation with the priming unit. In borrelidin biosynthesis, Modules BorM3, M4 and BorM5 utilize methylmalonyl-CoA. (175)

A short hydrophobic pocket in BorB permits various aliphatic substrates while preventing hydrolysis of diacid substrates. Like the fellow editing TEII, RifR, the pocket formed by the lid helices contains numerous charged residues; this may be a hallmark of promiscuous editing TE. (229) α/β hydrolases such as esterases or lipases are among the successfully engineered proteins and are often used for validating high-throughput enzyme engineering techniques. (247) However, to our knowledge, this success has not been translated to PKS and NRPS TEs. Lower throughput methods, such as rational mutagenesis of the substrate pockets of PKS and NRPS TEs often leads to diminution of activity. (159, 230, 248) In lieu of a directed evolution scheme and due to the unlikely

chance of rationally engineering point mutants, we opted to test chimeric mutants. Chimeric proteins have had some success in alter product profiles in the related TE14 family of hotdog fold TE. (249) However, the chimeras made in this study lead to a general reduction in activity. Mechanistically, this may have been mediated through differential conformational flexibility rather than direct substrate contact; Lid Loop 1 is distal from the binding site and is linked to binding-induced conformational changes in RifR. (229)

Here we provide a kinetic and structural characterization of BorB. While BorB may not be of utility for releasing the medium and long chain fatty acids, it may yet serve as an editing thioesterase. Hitherto, TEIIs have not been applied in an editing role in engineered assembly lines, despite their well recorded utility in Natural Product Biosynthesis. (207) BorB joins the cohort of kinetically and structurally characterized TEIs and TEIIs. (153, 207) A large library of sequence-structure-function data is essential for predicting TE activity *in silico*, a feat that is currently not possible.(153, 250)

4.5. Methods

4.5.1. BorB accession numbers

BorB can be accessed via Uniprot (accession #Q70I09) and Genbank (accession #CAE45660.1) the Protein DataBank will be available when this material is published in article format.

4.5.2. TEII sequence scraping, filtering and alignment.

ClusterTools was used to scrape the Minimum Information about a Biosynthetic Gene Cluster (MiBIG) database using BorB as an HMM seed and KS/condensation domains as constraints. (251, 252) The search returned several 243 sequences and contained several well-characterized TEIIs, such as the surfactin, rifamycin, and prodigiosin TEs. (229, 230, 239) Sequences over 300aa in length were discarded as they likely corresponded to *cis*-acting TEIs in full modules. The remaining 202 sequences were aligned using Clustal Omega. (253) The accession numbers are available in Supporting Table 4-1.

Supporting Table 4-1. Accession numbers for TEIIs

Accession	Name	Accession	Name	Accession	Name	Accession	Name
>ACG60780.1	acyltransferase	>ADO85582.1	TEII	>BAC87911.1	probable	>ADJ63839.1	thioesterase
>ANC94977.1	AlmM	>ADZ24985.1	TEII	>AAD24880.1	putative	>ADM46371.1	thioesterase
>ADI58654.1	AsuC15	>AEC14350.1	TEII	>AAK06795.1	putative	>AEC13083.1	thioesterase
>ABY83141.1	Azi2	>AEH42494.1	TEII	>AAM33656.1	putative	>AEP40917.1	thioesterase
>ABY83171.1	Azi33	>AEK75515.1	TEII	>AAQ90178.1	putative	>AFI57003.1	thioesterase
>ABY83143.1	Azi4	>AFP87526.1	TEII	>AAS79448.1	putative	>AFL48524.1	Thioesterase
>AAD21211.1	BacT	>AFR69340.1	TEII	>ACB37753.1	putative	>AFU82629.1	thioesterase
>ADC79627.1	BafH	>AGD80621.1	TEII	>ACG60736.1	putative	>AFV52205.1	thioesterase
>AAN32977.1	BarC	>AGZ15471.1	TEII	>ADU85986.1	putative	>AGI99484.1	thioesterase
>ALG65316.1	Cal20	>AHB82050.1	TEII	>AET98912.1	putative	>AHA12093.1	thioesterase
>ALG65331.1	Cal5	>AHB82061.1	TEII	>AEU17900.1	putative	>AHH53515.1	thioesterase
>ABD14710.1	CesT	>AJO72744.1	TEII	>AEZ53955.1	putative	>AKA59442.1	thioesterase
>AAZ77688.1	ChlK	>BAF85835.1	TEII	>AEZ53965.1	putative	>ALA09374.1	thioesterase
>ABR67751.1	CmnH	>BAO66517.1	TEII	>AGZ15461.1	putative	>BAB69188.1	thioesterase
>EFE73309.1	conserved	>BAQ25498.1	TEII	>AGZ15477.1	putative	>BAC68117.1	thioesterase
>ADI59536.1	CorM	>BAT51068.1	TEII	>BAC76483.1	putative	>BAD38876.1	thioesterase
>ABV83231.1	CppE	>CAC37888.1	TEII	>BAE98152.1	putative	>BAD55608.1	thioesterase
>AFD30951.1	CrmJ	>CAL58678.1	TEII	>BAH04169.1	putative	>BAF50718.1	thioesterase
>AKP45393.1	CysE	>CAQ52621.1	TEII	>BAI63286.1	putative	>BAO84863.1	thioesterase
>ABP57754.1	DepJ	>CBA11581.1	TEII	>CAC17504.1	putative	>BAP16685.1	thioesterase
>ABA59546.1	discrete	>CBW45643.1	TEII	>CAG69509.1	putative	>BAP34735.1	thioesterase
>AJW76718.1	DsaP	>CUI25709.1	TEII	>CAJ76283.1	putative	>CAA16185.1	thioesterase
>AHB38504.1	EpnB	>ACR50794.1	TEIIputative	>CAJ87593.1	putative	>CAE53380.1	thioesterase
>AFB35636.1	EsmB5	>AEA30276.1	TEIIputative	>CAJ88195.1	putative	>CAJ14052.1	thioesterase
>AAF86400.1	FkbQ	>AIL50171.1	TEIIputative	>CAK50771.1	putative	>CAJ14053.1	thioesterase
>AAQ82559.1	FscTE	>BAE93720.1	TEIIputative	>CAP20361.1	putative	>CAJ34372.1	thioesterase
>CBL93724.1	GetN	>BAG85019.1	TEIIputative	>CBW45730.1	putative	>CAJ34373.1	thioesterase
>EFG10342.1	Gramicidin	>BAH43770.1	TEIIputative	>CBZ42142.1	putative	>CAJ96475.1	Thioesterase
>CBK62748.1	holC	>BAH43865.1	TEIIputative	>CBZ42144.1	putative	>CAQ64680.1	thioesterase
>AJI44178.1	hypothetical	>BAQ21958.1	TEIIputative	>CDG12863.1	putative	>CBJ89763.1	Thioesterase
>CAI94722.1	hypothetical	>CAD91231.1	TEIIputative	>EJK79839.1	putative	>CBL93712.1	Thioesterase
>WP_052390330.1	hypothetical	>CCC55924.1	TEIIputative	>AAG52991.1	RifR	>CBW45734.1	thioesterase
>CAA73129.1	Irp4	>AAG02346.1	thioesterase	>CQR60477.1	CmxR	>CCB53268.1	Thioesterase
>ABV56586.1	KtzF	>AAN85504.1	thioesterase	>ADX66462.1	ScnI	>CCC21119.1	thioesterase
>ABX71117.1	Lct34	>AAN85542.1	thioesterase	>AAL15591.1	Sim13	>CUI25720.1	Thioesterase
>ABX71151.1	Lcz34	>AAO65810.1	thioesterase	>CAE02634.1	SrfAD	>CUI25737.1	Thioesterase
>AAU39362.1	lichenysin	>AAS79476.1	thioesterase	>ADC79640.1	TamB	>CUW01199.1	thioesterase

>CAQ71825.1	Linear	>AAU93800.1	thioesterase	>AAC69333.1	TEII	>EFG10349.1	Thioesterase
>ABB05106.1	LipTe	>AAX98200.1	thioesterase	>AAG13923.1	TEII	>ERG36781.1	thioesterase
>AGC09497.1	LobB2	>AAY93358.1	thioesterase	>AAN85527.1	TEII	>WP_010369421.1	thioesterase
>AHF22856.1	MarJ	>AAZ94405.1	thioesterase	>AAP92498.1	TEII	>WP_030890478.1	thioesterase
>CAD29792.1	microcystin	>ABI91471.1	Thioesterase	>AAU93793.1	TEII	>CAG15042.1	thioesterase
>WP_063834950.1	MULTISPECIES:	>ABO15862.1	thioesterase	>ABB52527.1	TEII	>CAM00070.1	thioesterase
>AAF71777.1	NysE	>ABP55205.1	Thioesterase	>ABC35463.1	TEII	>WP_043555673.1	thioesterase
>ABV97174.1	Oleoyl(acylcarrierp	>ABV56601.1	thioesterase	>ABC39254.1	TEII	>AGN74877.1	thioestrerase
>CAC20922.1	PimI	>ABW00324.1	Thioesterase	>ACN69977.1	TEII	>ALJ49917.1	TImA
>AAO65795.1	PKS	>ABX37388.1	Thioesterase	>ACO94474.1	TEII	>CAQ34930.1	TPA:
>AAQ84143.1	Plm8	>ACB46473.1	thioesterase	>ACO94502.1	TEII	>AHF21225.1	TriA
>AGC24272.1	PrlR	>ACR13893.1	thioesterase	>ACZ65480.1	TEII	>ALJ49928.1	TtmA
>ABV91293.1	probable	>ACR33073.1	thioesterase	>ADO85582.1	TEII	>ACM79811.1	ZmaP
>ABW96552.1	probable	>ACY06298.1	thioesterase				

4.5.3. Plasmid Construction

Plasmids were designed via the J5 software package. (210) The coding sequences and backbones were prepared via PCR amplification with Q5 polymerase (NEB). Templates were degraded via Dpn1 digestion (Thermo) and the correct bands were gel extracted and assembled via Gibson assembly. (214) Assemblies were transformed into *E. coli* DH10b and Sanger sequenced to verify inserts (Quintara). Plasmids used in this study are summarized Supporting Table 4-2. Physical samples and sequence maps, including primers, are available through the JBEI public registry. (<https://registry.jbei.org/>)

Supporting Table 4-2. Plasmids.

Plasmid	Source	Accession
pMAB01 (pET28a-6xHis-BorB)	This study	https://registry.jbei.org/entry/99510
pMAB02 (pBbS2K-6xHis-MBP-BorB)	This study	https://registry.jbei.org/entry/99511
pMAB03 (pBbS2K-6xHis-MBP-BorB-S96A)	This study	https://registry.jbei.org/entry/99512
pSC013 (pET28a-6xHis-BorM5-ACP)	This study	https://registry.jbei.org/entry/101737
pRK793 (pMAL-C2 TEV_protease)	David Waugh	https://www.addgene.org/8827/
pSC051 (pBbS2K-6xHis-MBP-BorB-RifR-Loop)	This study	https://registry.jbei.org/entry/101738
pSC052 (pBbS2K-6xHis-MBP-BorB-RedJ-Loop)	This study	https://registry.jbei.org/entry/101739
pSC053 (pBbS2K-6xHis-MBP-BorB-ScoT-Loop)	This study	https://registry.jbei.org/entry/101740
pSC049 (pET28a-6xHis-TEV-BorM5-ACP)	This study	https://registry.jbei.org/entry/99508
pARH130 (pBbS2k-6xHis-MBP-Borm5)	Curran et. al 2018.	https://registry.jbei.org/entry/33363

pMAB01 (pET28a-6xHis-BorB) and pMAB02 (pBbS2K-6xHis-MBP-BorB)

The open reading frame for BorB was codon optimized for *E. coli* and synthesized (Genewiz). BorB and backbone fragments were prepared by PCR and assembled with Gibson assembly. (40)

pMAB03 (pBbS2K-6xHis-MBP-BorB-S96A)

The S96A mutation was installed via overlapping PCR primers. The construct was split into two fragments via PCR and assembled with Gibson assembly. (40)

BorB chimeras (pSC051-53)

pMAB02 (6xHis-MBP-BorB) was linearized by double restriction digest with KpnI/NdeI (Thermo) and the fragment was gel extracted. A 450bp sequence corresponding to the excised KpnI/NdeI fragment was synthesized (Twist Bioscience) and included codon-optimized homologous sequences from RifR, RedJ, and ScoT. The linearized backbone and synthetic fragments were assembled via Gibson assembly, transformed into *E. coli* DH10b and Sanger sequenced to verify inserts (Quintara).

BorM5-ACP

Codon-optimized BorM5-ACP and pET28a were PCR amplified, assembled by Gibson assembly, transformed into *E. coli* DH10b and Sanger sequenced to verify inserts (Quintara). The codon sequence for BorM5-ACP was derived from pARH130. (254)

4.5.4. Protein purification for biochemical assays

6xHis-MBP-TEV-BorB and *6xHis-MBP-TEV-BorB-S96A*

BL-21 (DE3) cells containing expression plasmids were cultured overnight at 37°C, 200 rpm in 50 mL Terrific Broth (TB), 50 µg*mL⁻¹ kanamycin before being inoculated 1% v/v into 2 L TB media, 50 µg*mL⁻¹ kanamycin. Cells were split into four 2 L baffled shake flasks and incubated at 37°C, 200 rpm until they reached OD₆₀₀ ≥ .2 AU. Protein expression was induced with 50 ng*mL⁻¹ anhydrotetracycline and cells were incubated at 18°C, 200 rpm for 20 hours. Cells were centrifuged (6000 × g, 10 min, 4°C), the supernatant was discarded, and the cells were suspended in 50 mL lysis buffer (50 mM NaPO₄²⁻, pH 7.5, 150 mM NaCl, 20 mM imidazole). Cells were stirred at 200 rpm on ice until clumps were no longer visible (15-20 minutes). Cells were lysed by 5-10 passes through an Avestin Emulsiflex C3 homogenizer (15,000 bar) and centrifuged (40,000 × g, 30 min, 4°C) to pellet the insoluble fraction.

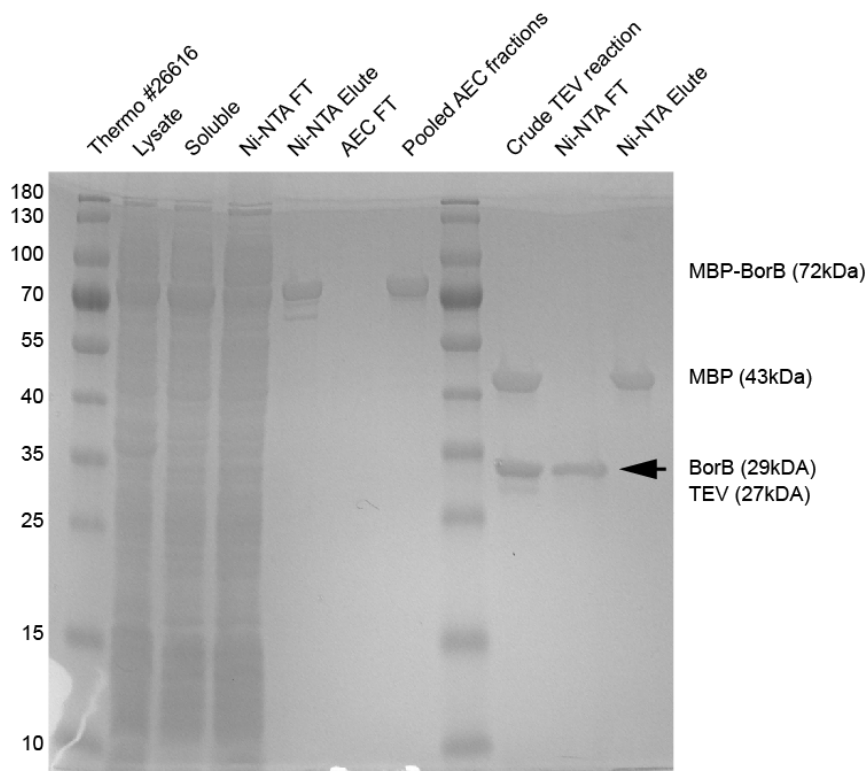
NiNTA purification was done using 5 mL GE HisTrap column on an AktA FPLC. The column was equilibrated in 10 CV lysis buffer, 10 CV elution buffer (50 mM NaPO₄²⁻, pH 7.5, 150 mM NaCl, 500 mM imidazole), and 10 CV lysis buffer. The soluble fraction of the lysate was loaded onto the column via the sample pump at 2 mL*min⁻¹ and the FT was collected. The column was washed with 10 CV lysis buffer. The bound protein was eluted from 20-500 mM imidazole at 2.5 mL*min⁻¹ over 100 mL. The desired fractions were identified by SDS-PAGE.

The eluent was dialyzed to remove the imidazole and salt. ~20 mL protein from NiNTA was incubated in 1 L AEC buffer A (50 mM NaPO₄²⁻, pH 7.5, 50 mM NaCl, 1 mM DTT) for 1 hour at 4°C. This was repeated. AEC purification was done using a 5 mL HiTrap Q anion column on an AktA FPLC. The column was equilibrated with 10 CV AEC buffer A (50 mM NaPO₄²⁻, pH 7.5, 50 mM NaCl, 1 mM DTT), 10 CV AEC buffer B (50 mM

NaPO₄²⁻, pH 7.5, 1 M NaCl, 1 mM DTT) and an additional 10 CV buffer A. The dialyzed protein sample was loaded via the sample pump at 5 mL*min⁻¹. The bound protein was washed with 10CV buffer A. The bound protein was eluted in a gradient (0-100% buffer B over 60 minutes at 1 mL*min⁻¹). The desired fractions were identified by SDS-PAGE.

The protein concentration in the pooled AEC fractions was quantified by UV absorbance. The sample was diluted in AEC buffer A to ~50 mM NaCl and ~2 mg*ml⁻¹ protein. TEV protease was added at 1:50 molar ratio. The cleavage reaction proceeded for 16 hours at 4°C and was verified by SDS-PAGE. Subtractive NiNTA was performed as stated above with modifications. The reaction was run through an equilibrated NiNTA column at 1 mL*min⁻¹ and washed with an additional ~20 mL of AEC buffer A. The flowthrough (cleaved protein) and eluent (tag, TEV protease, residual un-cleaved protein) were collected. The purity of the flowthrough was verified by SDS-PAGE (Supporting Figure 4-1). The flowthrough was concentrated and buffer exchanged by spin filtration into 50 mM NaPO₄²⁻, pH 7.5, 8% glycerol, 1 mM DTT, flash frozen in liquid nitrogen, and stored at -80°C.

6xHis-BorM5-ACP



Supporting Figure 4-1. Purification and TEV-cleavage of BorB. Arrow points to BorB used for biochemical reactions.

The purification for 6xHis-BorM5-ACP followed the steps laid out for the BorB proteins, with the exception of the TEV-mediated cleavage.

TEV protease

The purification for the TEV protease followed the original published protocol. (211)

SFP

Purified SFP was kindly provided by Jacquelyn Blake-Hedges.

4.5.5. Protein expression for crystallography

BL-21 (DE3) cells containing expression pMAB02 were pre-cultured overnight at 37°C, 200 rpm in 50 mL Terrific Broth (TB), 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin before being inoculated 1% v/v into 2 L TB media, 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin. Cells were split into four 2 L baffled shake flasks and incubated at 37°C, 200 rpm until they reached $\text{OD}_{600} \geq .2$ AU. Protein expression was induced with 50 $\text{ng}\cdot\text{mL}^{-1}$ anhydrotetracycline and cells were incubated at 18°C, 200 rpm for 20 hours.

Cells were centrifuged (6000 x g, 10 min, 4°C), the supernatant was discarded, and the cells were suspended in 50 mL lysis buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 20 mM imidazole). Cells were stirred at 200 rpm on ice until clumps were no longer visible (15-20 minutes). Cells were lysed by 5-10 passes through an Avestin Emulsiflex C3 homogenizer (15,000 bar) and centrifuged (40,000 x g, 30 min, 4°C) to pellet the insoluble fraction.

NiNTA purification was done using 5 mL GE HisTrap column on an AktA FPLC. The column was equilibrated in 10CV lysis buffer, 10CV elution buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 500 mM imidazole), and 10CV lysis buffer. The soluble fraction of the lysate was loaded onto the column via the sample pump at $2\text{mL}\cdot\text{min}^{-1}$ and the FT was collected. The column was washed with 10CV lysis buffer. The bound protein was eluted at $2.5\text{mL}\cdot\text{min}^{-1}$ over 100 mL. Fractions were analyzed by SDS-PAGE.

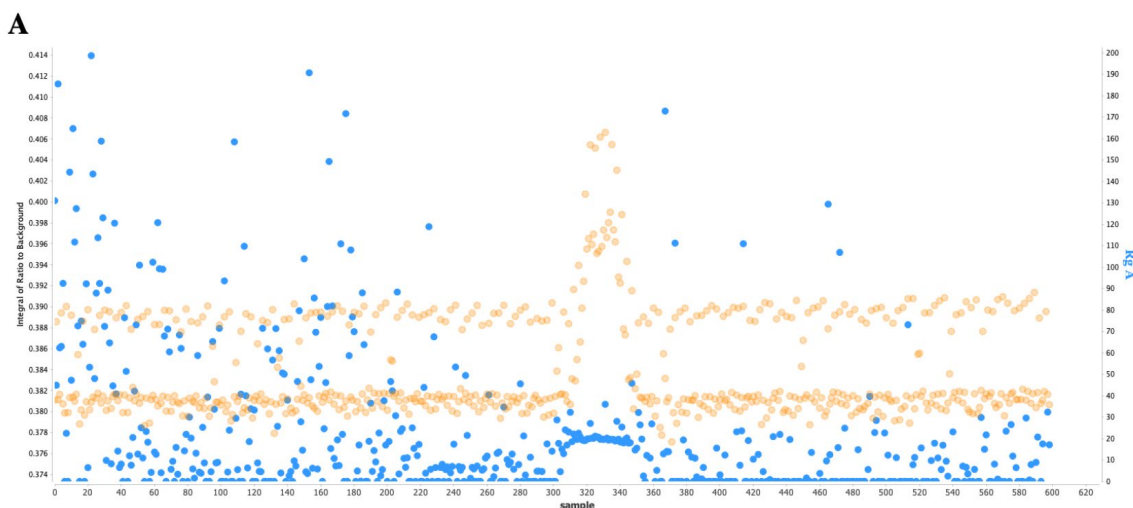
Proteins were dialyzed to remove the imidazole and salt. ~20 mL protein from NiNTA was incubated in 1L AEC buffer A (25 mM HEPES, pH 7.5, 50 mM NaCl, 1mM DTT) for 1hr at 4°C. This was repeated.

AEC purification was done using a 5 mL HiTrap Q anion column on an AktA FPLC. The column was equilibrated in 10CV AEC buffer A (25 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM DTT) 10CV AEC buffer B (25 mM HEPES, pH 7.5, 1M NaCl, 1 mM DTT) and 10CV buffer A. The dialyzed protein sample was loaded via the sample pump at $5\text{mL}\cdot\text{min}^{-1}$. The bound protein was washed with 10CV buffer A. The bound protein was eluted in a gradient (0-100% buffer B over 60 minutes at $1\text{mL}\cdot\text{min}^{-1}$). The desired fractions were identified by SDS-PAGE.

The protein in pooled AEC fractions was quantified. The sample was dialyzed in buffer A to ~50 mM NaCl and ~2mg*ml⁻¹ protein. TEV protease was added at 1:100 molar ratio. The cleavage reaction proceeded for 16 hours at 4°C and was verified by SDS-PAGE. Subtractive NiNTA was performed as stated above with modifications. The reaction was run through an equilibrated NiNTA column at 1 mL*min⁻¹ and washed with an additional ~20 mL of AEC buffer A. The flowthrough (cleaved protein) and eluent (tag, TEV protease, residual un-cleaved protein) were collected. The sample was concentrated to 10 mg*mL⁻¹ and distributed in crystallization trays immediately.

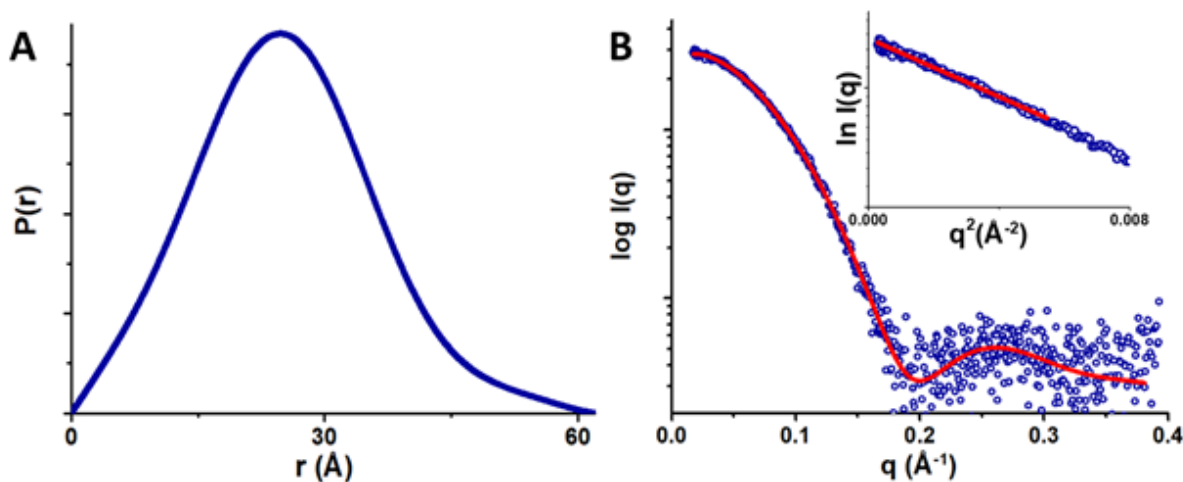
4.5.6. Size-Exclusion Chromatography-Small Angle X-ray Scattering (SEC-SAXS)

Prior to SEC-SAXS experiments, the samples were prepared as stated in section 4.5.5. The final concentration of cleaved BorB was 10 mg*mL⁻¹. SEC-SAXS profile was collected at the ALS beamline 12.3.1 LBNL Berkeley, California. (255) X-ray wavelength $\lambda = 1.127 \text{ \AA}$ and the sample-to-detector distances were set to 2105 mm resulting in scattering vectors, q , ranging from 0.01 \AA^{-1} to 0.4 \AA^{-1} . The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20°C (256) and data was processed as described. (257) Briefly, the flow through SAXS cell was directly coupled with an online Agilent 1260 Infinity HPLC system using a Shodex KW-802.5 column. The column was equilibrated with running buffer (25 mM HEPES, 50 mM NaCl, 1 mM DTT, 2% glycerol, pH 7.5) with a flow rate of 0.55 mL/min. The 65 μL sample was run through the SEC and 3 second X-ray exposures were collected continuously during a 30 min elution. The SAXS frames recorded prior to the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by radius of gyration R_g



Supporting Figure 4-2. SEC-SAXS Signal Plot. Each point represents the integrated area of the ratio of the sample SAXS curve to the estimated background.

derived by the Guinier approximation $I(q) = I(0) \exp(-q^2 R_G^2/3)$ with the limits $qR_G < 1.3$. Integral of ratios to background and R_G values were compared for each collected SAXS curve (frame) across the entire elution peak using the program SCÅTTER. The elution peak was mapped by plotting the integral of ratios to background and R_G relative to the recorded frame (Supporting Figure 4-2). Uniform R_G values across an elution peak represent a homogenous sample. The merged SAXS profile was additionally investigated for aggregation using Guinier plots (Supporting Figure 4-3). The program SCÅTTER was used to compute the pair distribution function $P(r)$ (Supporting Figure 4-3). The distance r where $P(r)$ approach zero intensity identifies the maximal dimension of the macromolecule (D_{max}). $P(r)$ functions were normalized based on the molecular weight of the assemblies as determined by SCÅTTER using volume of correlation V_c .(258) These SAXS profiles were then compared to theoretical scattering curves generated from atomistic models using the FOXS.(237, 238)



Supporting Figure 4-3. SAXS analyses of BorB. A) Normalized $P(r)$ function calculated for the experimental data and showing the maximal dimension of the macromolecule (D_{max}) B) SAXS profile for BorB (blue) together with the fit for $P(r)$ function. Inset, Guinier plots for the SAXS curves.

Supporting Table 4-3. SAXS parameters.

	MW_{monomer} (kDa)	MW_{SAXS} (kDa)	R_g (Å)	D_{max} (Å)
BorB	29	23	19	64

4.5.7. Crystallization, X-ray data collection and structure determination

The final concentration of BorB used for crystallization trials was 10 mg*ml⁻¹. The BorB protein was screened using the sparse matrix method (259) with a Phoenix Robot (Art Robbins Instruments, Sunnyvale, CA) using the following crystallization screens: Berkeley Screen, (260) Crystal Screen, SaltRx, PEG/Ion, Index and PEGRx (Hampton Research, Aliso Viejo, CA). Crystals of BorB were found in 0.2 M Magnesium Chloride, 0.1 M HEPES pH 7.5 and 22 % Poly (acrylic acid sodium salt) 5,100. BorB crystals were obtained by the sitting-drop vapor-diffusion method with the drops consisting of a mixture of 0.2 µl of protein solution and 0.2 µl of reservoir solution.

A crystal of BorB was placed in a reservoir solution containing 20% (v/v) glycerol, then flash-cooled in liquid nitrogen. The X-ray data sets for BorB were collected at the Berkeley Center for Structural Biology beamline 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory (LBNL). The diffraction data were recorded using an ADSC-Q315r detector. The data sets were processed using the program Xia2. (261)

The BorB crystal structure was determined by the molecular-replacement method with the program PHASER (262) within the Phenix suite, (263, 264) using as a search model the structure RifR, the Type II Thioesterase (PDB code 3FLA), ((229) which has 50 % sequence identity to the target. The atomic positions obtained from molecular replacement and the resulting electron density maps were used to build the BorB structure and initiate crystallographic refinement and model rebuilding. Structure refinement was performed using the phenix.refine program. (2) Translation-libration-screw (TLS) refinement was used, with each protein chain assigned to a separate TLS group. Manual rebuilding using COOT (265) and the addition of water molecules allowed construction of the final model. The final model of BorB has an R factor of 23.3 % and a R_{free} of 26.9 %. Root-mean-square deviation differences from ideal geometries for bond lengths, angles and dihedrals were calculated with Phenix. (1) The stereochemical quality of the final model of BorB was assessed by the program MOLPROBITY. (266) Crystallization parameters are listed in Supporting Table 4-4.

Supporting Table 4-4. Statistics for X-rays data collection and structure refinement.

A. Data collection	
Wavelength (Å)	0.99998
Resolution range (Å)	46.23 - 1.897 (1.965 - 1.897)
Detector Distance (mm)	220
Φ (deg.) collected / ΔΦ (deg.)	200 / 1.0
Exposure time (seconds)	3
Temperature of collect (Kelvin)	100
B. Data statistics	
Space group	C 2 2 2 ₁
Unit-Cell parameters (Å)	a=72.52, b=92.46 and c=150.21
Total reflections	318670 (30129)
Unique reflections	39997 (3844)
Multiplicity	8.0 (7.9)
Data completeness (%)	95.33 (80.49)d
Mean I/ sigma (I)	10.07 (1.60)
Wilson B-factor	20.17
R-merge	0.1439 (1.233)
R-meas	0.1538 (1.319)
R-pim	0.05394 (0.4631)
CC1/2	0.997 (0.874)
CC*	0.999 (0.966)
C. Structure Refinement	
Reflections used in refinement	38446 (3217)
Reflections used for R-free	1870 (132)
R-work	0.233 (0.6585)
R-free	0.269(0.6762)
CC (work)	0.950 (0.226)
CC(free)	0.904 (0.167)
RMS from ideal geometry	
Bond lengths (Å)	0.005
Bond angles (°)	1.10
Protein residues	480
Water molecules	269
Average isotropic B-factors	
Protein atoms (Å ²)	39.3
Solvent atoms (Å ²)	39.1
Ramachandran Plot	
Favored region (%)	97.2
Outliers region (%)	0.2

4.5.8. Manual substrate docking

Docking was performed visually with Maestro (Schrodinger).

4.5.9. NMR

NMR spectra were acquired at the UC Berkeley College of Chemistry NMR facility on Bruker Avance AVB-400 spectrometer. Data were processed using Bruker Topspin. Data were analyzed using MestReNova v14.0.0.

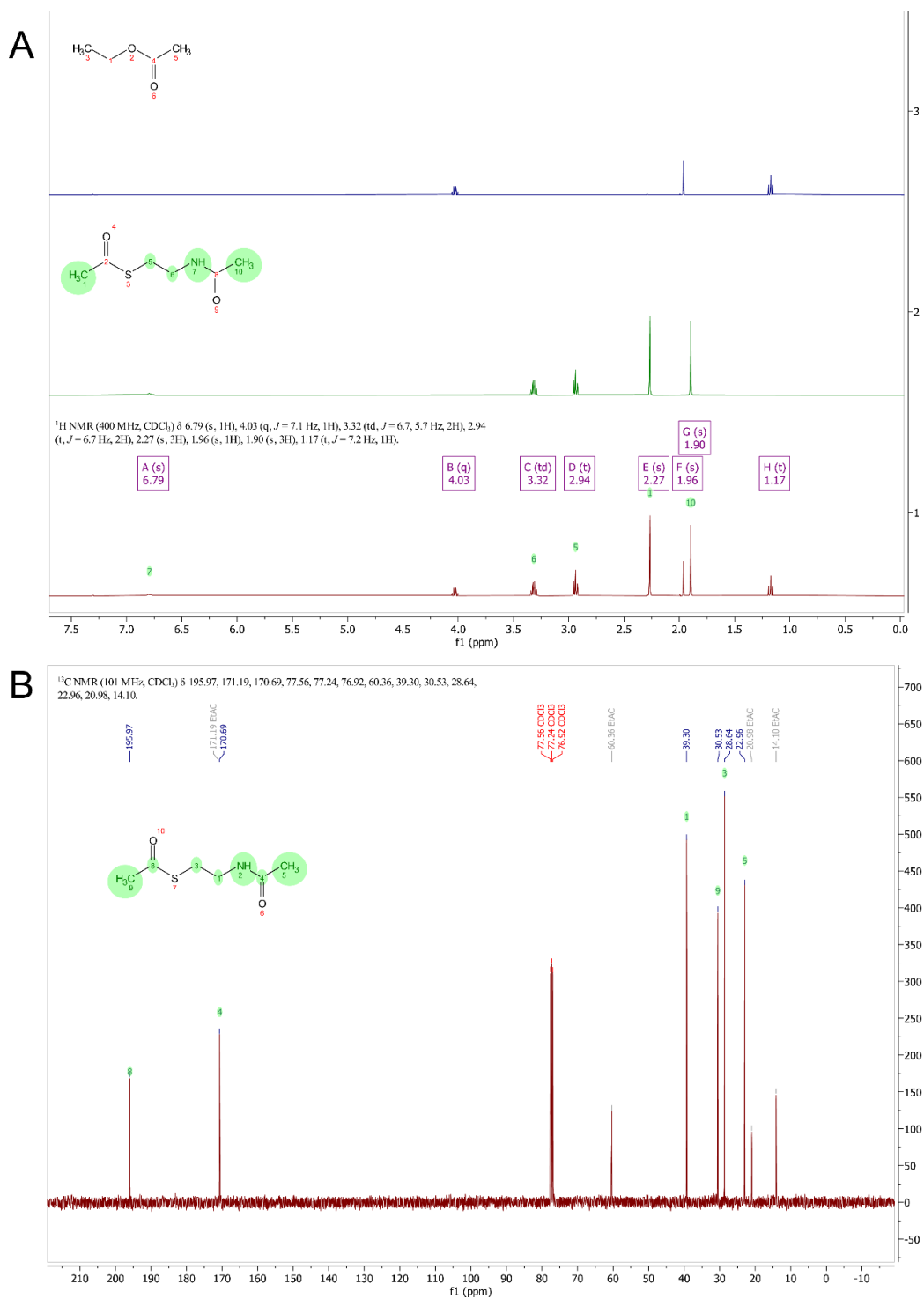
4.5.10. Acyl-SNAC synthesis

S-(2-acetamidoethyl) ethanethioate

To a cold solution of .1 M NaHCO₃, 1 equivalent (0.002 mols) of acetic anhydride was added into a round bottom flask. *N*-acetylcysteamine (1eq.) and triethylamine (1eq.) was then added dropwise, and the mixture was reacted overnight at room temperature. Product was extracted with DCM and concentrated using a roto-evaporator. Unreacted *N*-acetylcysteamine was removed using copper impregnated silica gel. Remaining impurities were removed using silica column purification (ethyl acetate:hexanes, increasing ratio over time). Yield: 138 mgs, 43%. The NMR spectrum was contaminated with residual ethyl acetate; this was evaporated with excessive rotary evaporation before use. For spectra see Supporting Figure 4-4.

¹H NMR (400 MHz, CDCl₃) δ 6.79 (s, 1H), 4.03 (q, *J* = 7.1 Hz, 1H), 3.32 (td, *J* = 6.7, 5.7 Hz, 2H), 2.94 (t, *J* = 6.7 Hz, 2H), 2.27 (s, 3H), 1.96 (s, 1H), 1.90 (s, 3H), 1.17 (t, *J* = 7.2 Hz, 1H).

¹³C NMR (101 MHz, CDCl₃) δ 195.97, 171.19, 170.69, 77.56, 77.24, 76.92, 60.36, 39.30, 30.53, 28.64, 22.96, 20.98, 14.10.



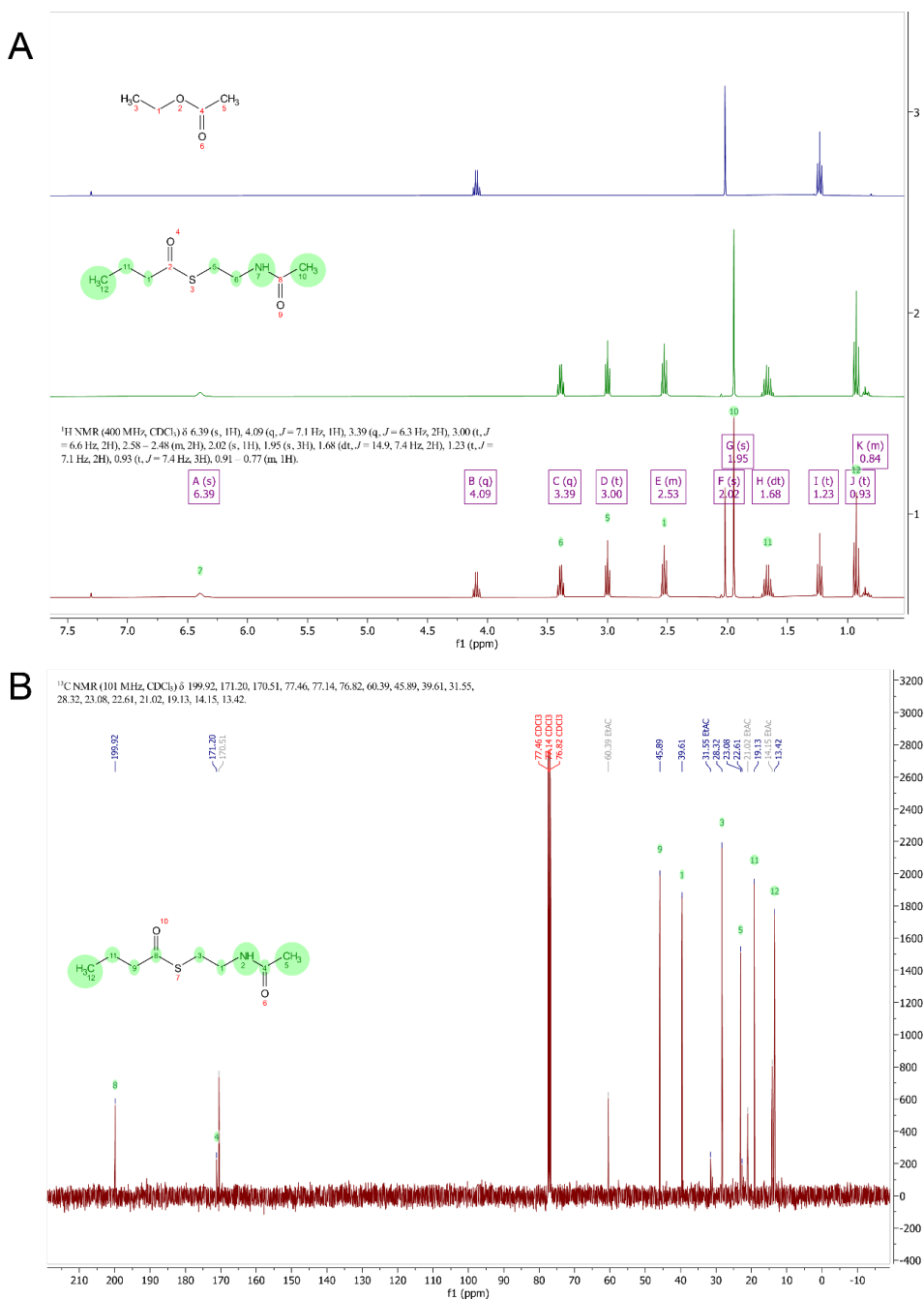
Supporting Figure 4-4. NMR spectra of S-(2-acetamidoethyl) ethanethioate. **A)** ¹H NMR. Bottom; full spectrum. Middle; peaks for S-(2-acetamidoethyl) ethanethioate. Top; peaks for ethyl acetate. **B)** ¹³C NMR. Assignments are in green.

S-(2-acetamidoethyl) butanethioate

To a cold solution of .1 M NaHCO₃, 1 equivalent (.002 mols) of butyric anhydride was added into a round bottom flask. *N*-acetylcysteamine (1eq.) and triethylamine (1eq.) was then added dropwise, and the mixture was reacted overnight at room temperature. Product was extracted with ether and concentrated using a roto-evaporator. Unreacted *N*-acetylcysteamine was removed using copper impregnated silica gel. Remaining impurities were removed using silica column purification (ethyl acetate:hexanes, increasing ratio over time). Yield 67 mgs, 18%. The NMR spectrum was contaminated with residual ethyl acetate; this was evaporated with excessive rotary evaporation before use. For spectra see Supporting Figure 4-5.

¹H NMR (400 MHz, CDCl₃) δ 6.39 (s, 1H), 4.09 (q, *J* = 7.1 Hz, 1H), 3.39 (q, *J* = 6.3 Hz, 2H), 3.00 (t, *J* = 6.6 Hz, 2H), 2.58 – 2.48 (m, 2H), 2.02 (s, 1H), 1.95 (s, 3H), 1.68 (dt, *J* = 14.9, 7.4 Hz, 2H), 1.23 (t, *J* = 7.1 Hz, 2H), 0.93 (t, *J* = 7.4 Hz, 3H), 0.91 – 0.77 (m, 1H).

¹³C NMR (101 MHz, CDCl₃) δ 199.92, 171.20, 170.51, 77.46, 77.14, 76.82, 60.39, 45.89, 39.61, 31.55, 28.32, 23.08, 22.61, 21.02, 19.13, 14.15, 13.42.



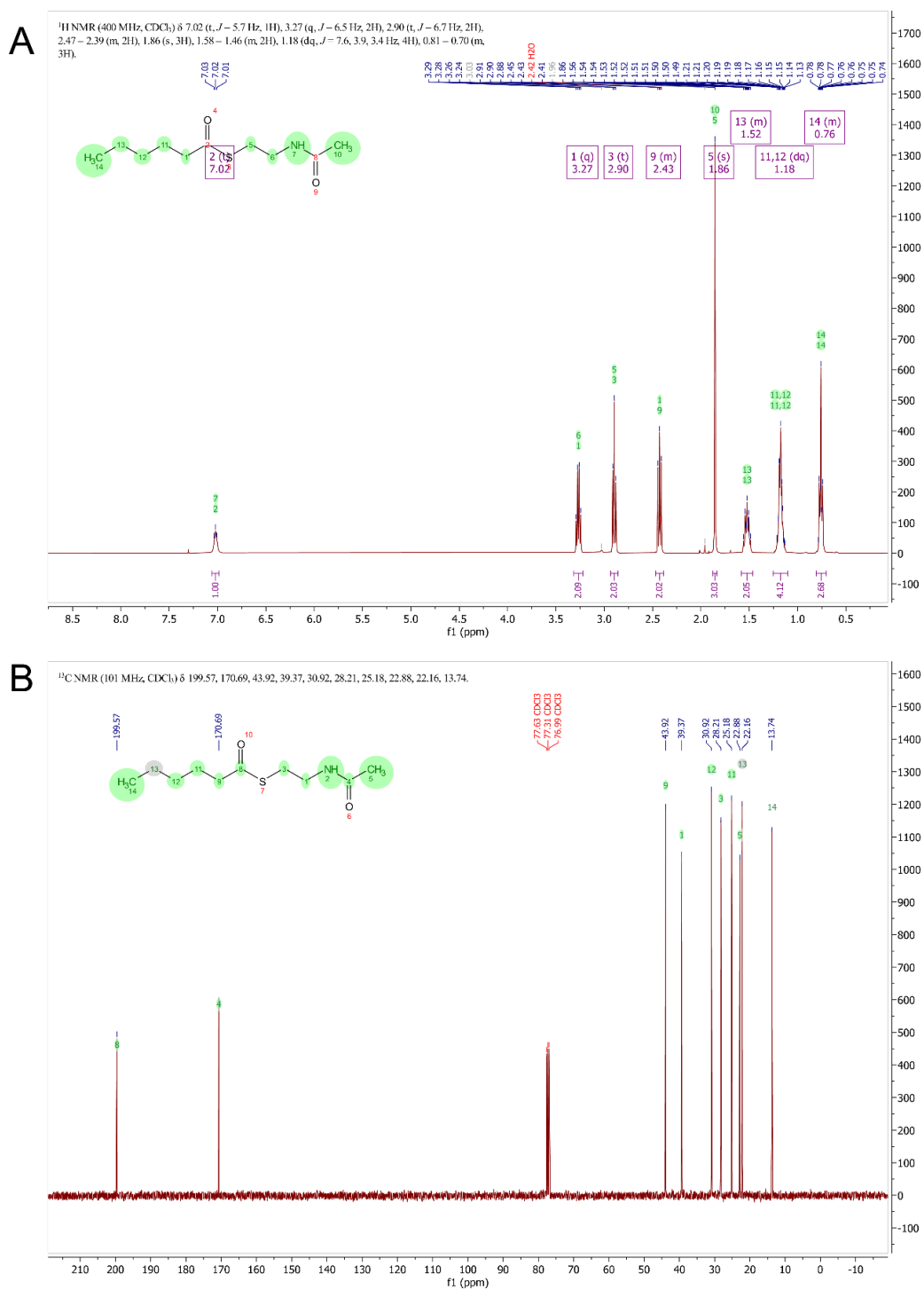
Supporting Figure 4-5. NMR spectra of S-(2-acetamidoethyl) butanethioate. A) ^1H NMR. Bottom; full spectrum. Middle; peaks for S-(2-acetamidoethyl) butanethioate. Top; peaks for ethyl acetate. B) ^{13}C NMR. Assignments are in green.

S-(2-acetamidoethyl) hexanethioate

To a cold solution of .1M NaHCO₃, 1 equivalent (.002 mols) of hexanoic anhydride was added into a round bottom flask. *N*-acetylcysteamine (1eq.) and triethylamine (1eq.) was then added dropwise, and the mixture was reacted overnight at room temperature. Product was extracted with ether and concentrated using a roto-evaporator. Unreacted HSNAC was removed using copper impregnated silica gel. Remaining impurities were removed using silica column purification (ethyl acetate:hexanes, increasing ratio over time). Yield 280 mgs, 64%. For spectra, see Supporting Figure 4-6.

¹H NMR (400 MHz, CDCl₃) δ 7.02 (t, *J* = 5.7 Hz, 1H), 3.27 (q, *J* = 6.5 Hz, 2H), 2.90 (t, *J* = 6.7 Hz, 2H), 2.47 – 2.39 (m, 2H), 1.86 (s, 3H), 1.58 – 1.46 (m, 2H), 1.18 (dq, *J* = 7.6, 3.9, 3.4 Hz, 4H), 0.81 – 0.70 (m, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 199.57, 170.69, 43.92, 39.37, 30.92, 28.21, 25.18, 22.88, 22.16, 13.74.



Supporting Figure 4-6. NMR spectra of S-(2-acetamidoethyl) hexananethioate. A) ^1H NMR. B) ^{13}C NMR. Assignments are in green.

4.5.11. Turnover assays

Turnover assays were performed in 200 mM phosphate buffer (pH 7.6), 0.2 mM 5,5-dithiobis-(2-nitrobenzoic acid) (DTNB) with 1 μ M BorB and substrate concentrations as stated in 100 μ L volumes. Assays were monitored continuously at 412 nm in a 96-well plate reader (SpectraMax). Absolute concentrations of H-SNAC were determined by comparison with an H-SNAC standard curve under the same conditions.

4.5.12. Kinetic Assays

Acyl-CoA kinetic assays were performed in 200 mM phosphate buffer (pH 7.6), 0.2 mM 5,5-dithiobis (2-nitrobenzoic acid) (DTNB). The final well volume was 20 μ L. Reactions were performed in well triplicate at 30°C in a 384 well clear bottom plate (Thermo 242764). Substrates were dispensed using the Echo 550 Liquid Handler. Reactions were initiated by the manual addition of 1 μ M purified BorB, or boiled (denatured) protein serving as a negative control, to the reaction mix. Spectrophotometric measurement was taken over 15 minutes, with absorbance at 412 nm being collected every 30 seconds. Free thiol concentrations were calculated by comparison to a standard curve of HSNAC. All wells were first background subtracted and then rates were corrected for non-enzymatic cleavage. The linear regions of activity were used to derive kinetic constants.

4.5.13. Kinetics derivatization

Under standard saturating conditions, the Michaelis-Menten equation follows Supporting Equation 4-1.

Supporting Equation 4-1. Michaelis-Menten Equation. V_o is initial velocity, V_{max} is maximum velocity, $[S]$ is substrate concentration, and K_m is the Michaelis-Menten constant.

$$V_o = \frac{V_{max} * [S]}{K_m + [S]}$$

At low substrate concentrations, i.e. when $K_m \gg [S]$, then $K_m + [S] \approx K_m$. Supporting Equation 4-2 formalizes this relationship.

Supporting Equation 4-2. Michaelis-Menten at low substrate concentrations.

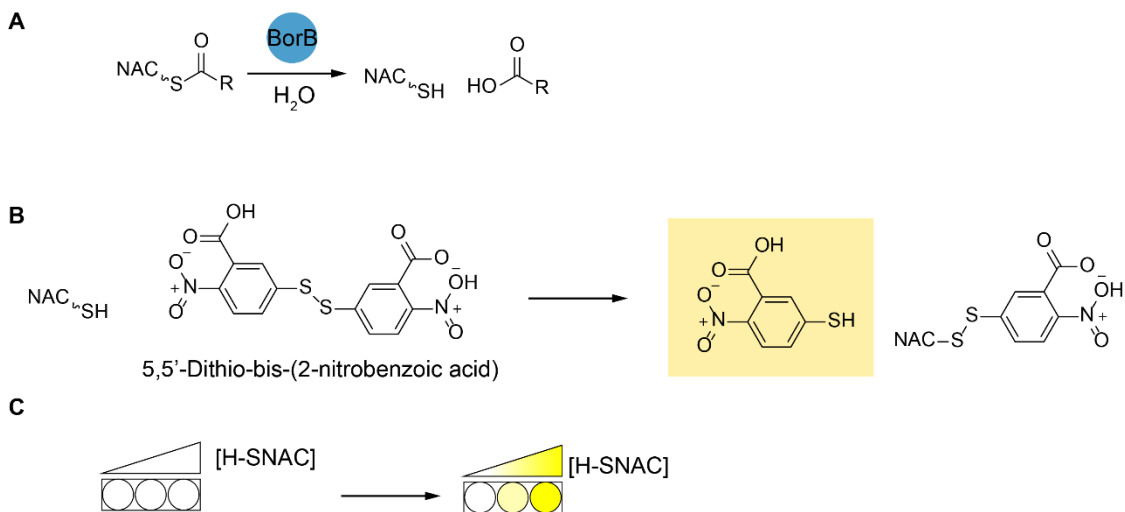
$$V_o = \frac{V_{max} * [S]}{K_m}$$

Under this regime, V_{max}/K_m is the slope of the plot of V_0 against $[S]$.

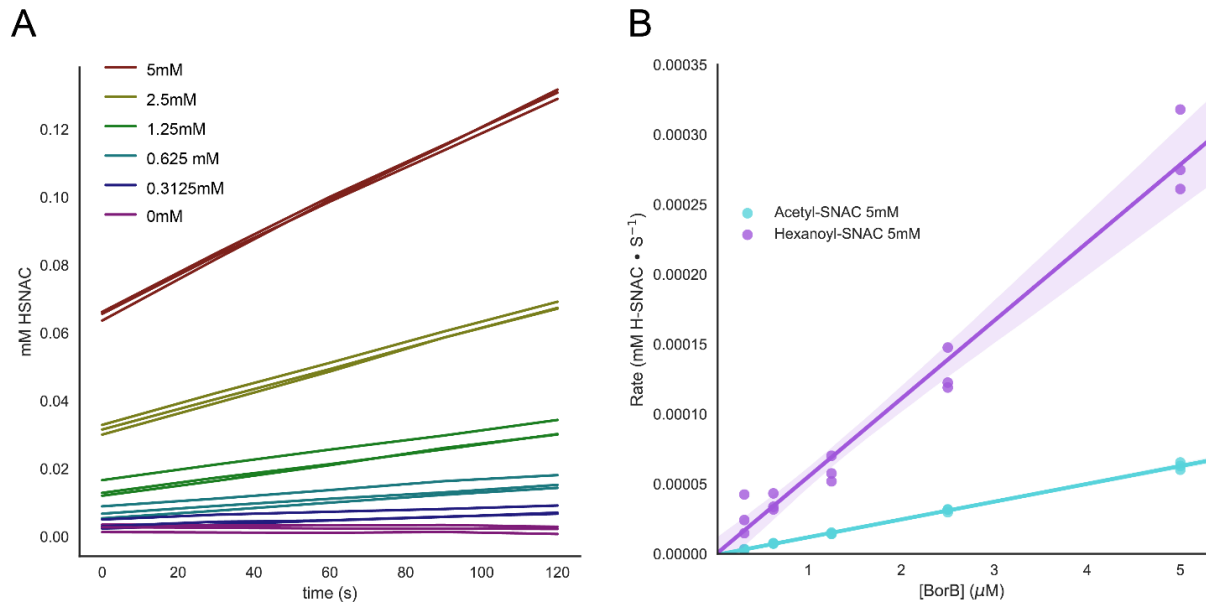
4.5.14. In vitro ACP-acylation and MALDI-TOF

50 μM *apo*-BorM5-ACP was mixed with 1 μM SFP and 250 μM acyl-CoA and incubated at 25°C for one hour. Samples were mixed in a 1:1 dilution series with saturated sinapinic acid in 30% H_2O , 70% acetonitrile and spotted onto a 384-well steel plate (Applied Biosystems). Samples were analyzed via an AB Sciex 4800 MALDI TOF under linear positive mode. Data were analyzed with AB Sciex Data Explorer.

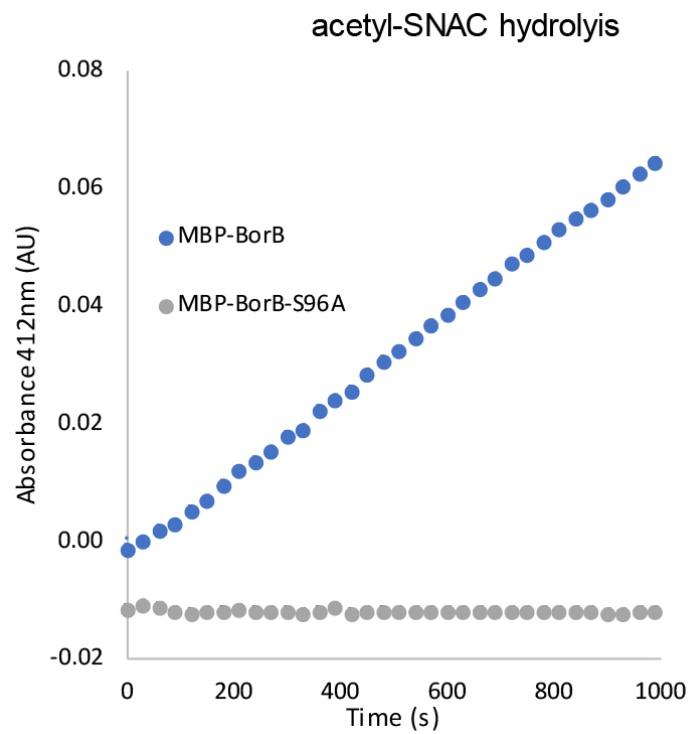
4.6. Supporting Figures



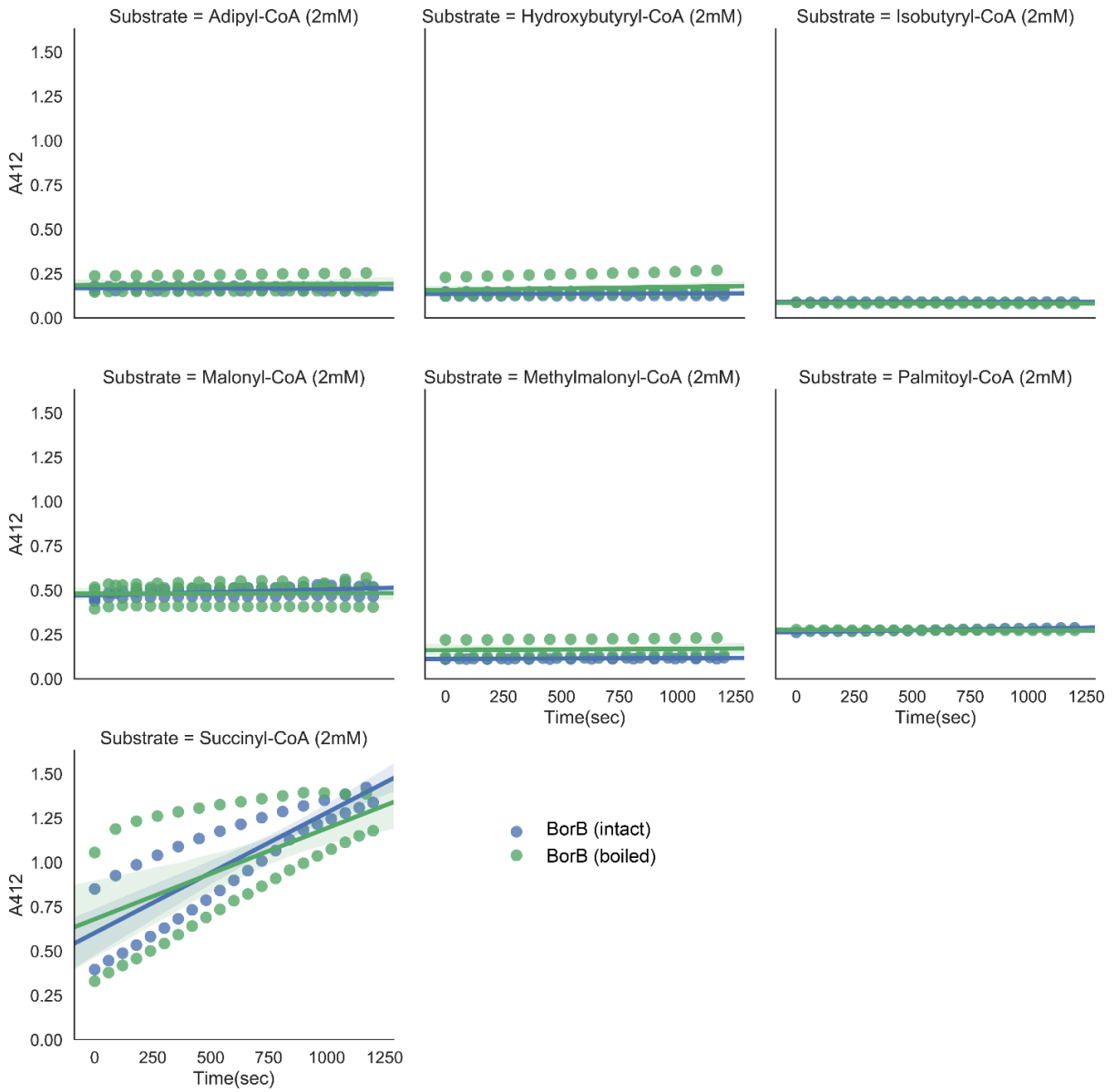
Supporting Figure 4-7. Ellman's assay. A) Thioester hydrolysis generates free fatty acids and thiols. B). Ellman's reagent (5,5'-dithio-bis-(2-nitrobenzoic acid)) is reduced by free thiols, yielding a yellow product. C) Thiols react quantitatively with DTNB, allowing quantitation over a linear range (0-0.5 mM).



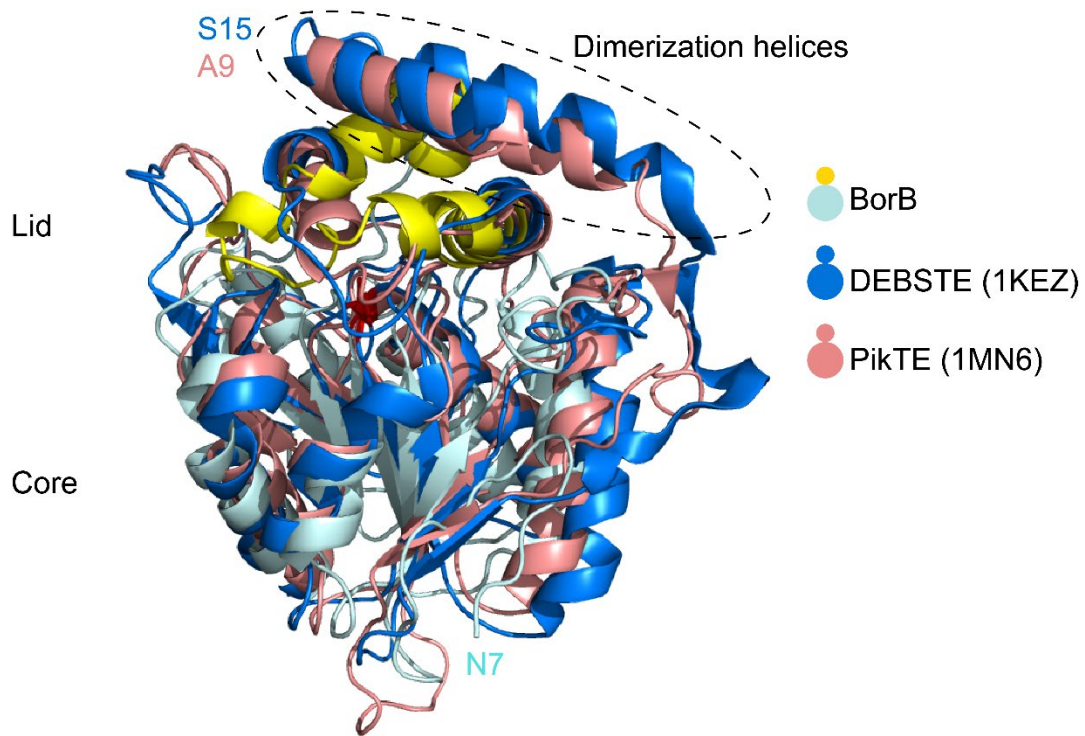
Supporting Figure 4-8. Acyl-SNAC hydrolysis assays. A) BorB-mediated cleavage of different concentrations of hexanoyl-SNAC. B) Rates of acyl-SNAC cleavage (5 mM) with varying concentrations of BorB.



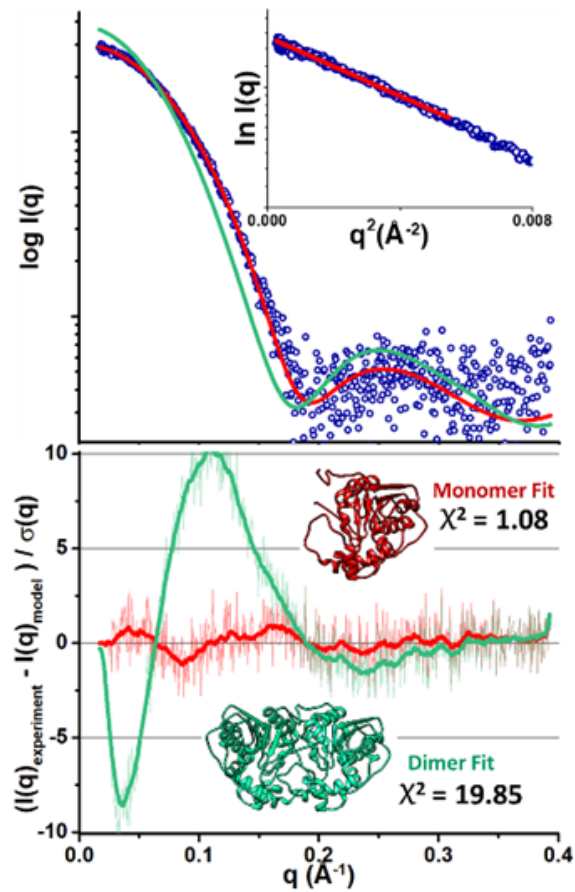
Supporting Figure 4-9. Activity of WT MBP-BorB (blue) and MBP-BorB Ser96Ala with 10 mM acetyl-SNAC.



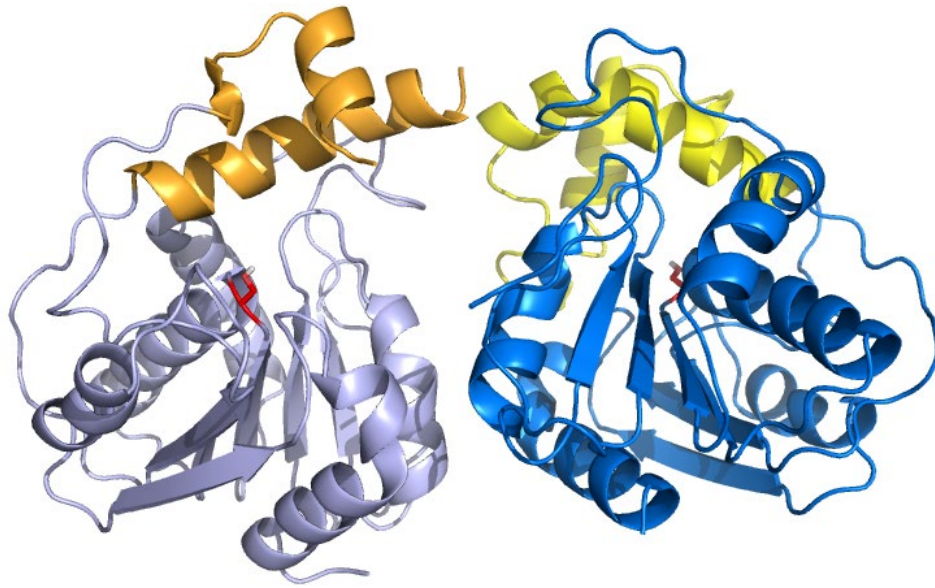
Supporting Figure 4-10. Hydrolysis of 2mM acyl CoA as indicated for each graph by boiled and intact BorB (1 μ M).



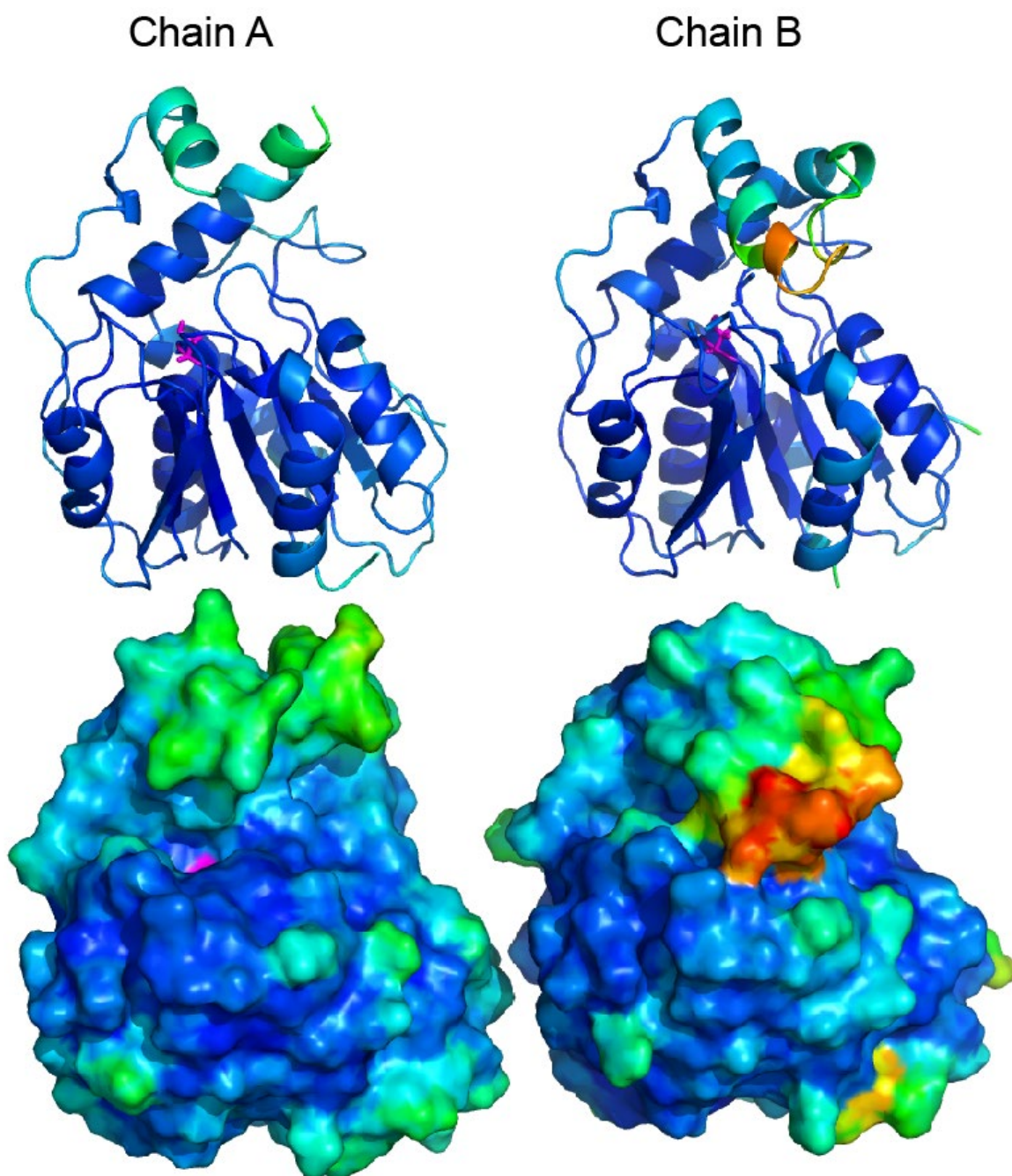
Supporting Figure 4-11. BorB (light blue and yellow) overlaid with structures of the DEBSTE (dark blue) and PikTE (salmon). The lid and active site serine of BorB are colored red for reference. The N-terminal helices of DEBSTE and PikTE are circled. The N-terminal residues are labeled for each construct.



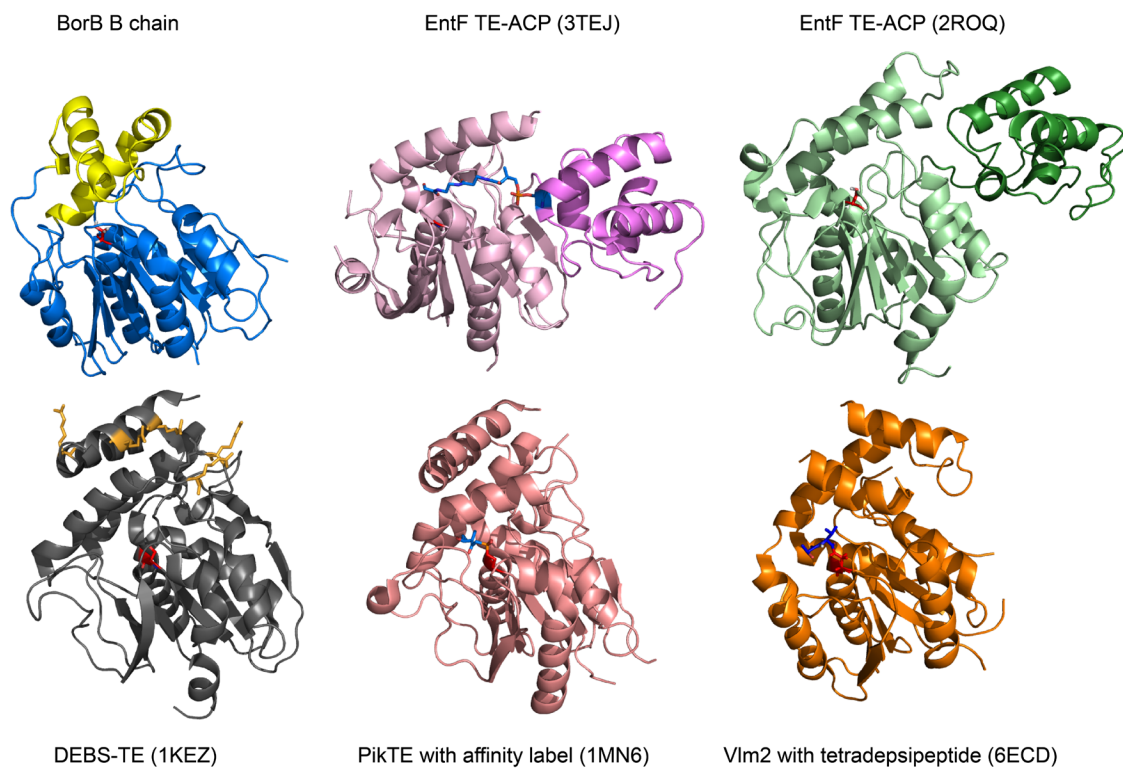
Supporting Figure 4-12. SAXS analyses of BorB. Top: SAXS profile for BorB together with the fits for monomer (red) and dimer (green). Inset, Guinier plots for the SAXS curves. Bottom: Residuals for SAXS profile fits, together with crystal structure and chi values of fits.



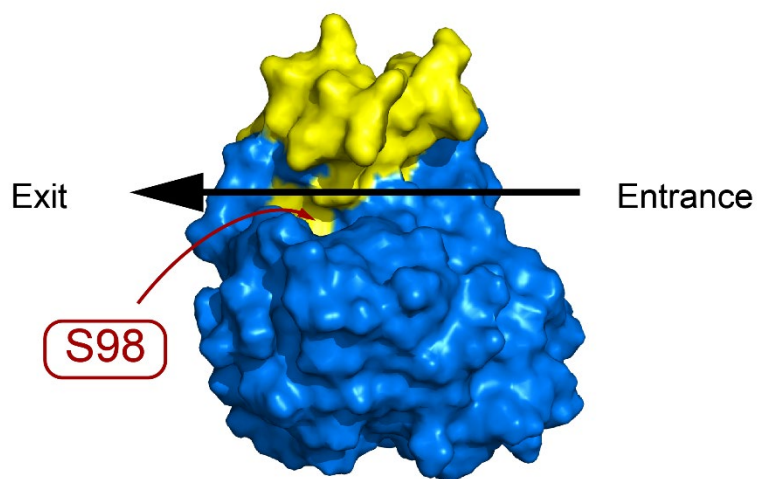
Supporting Figure 4-13. Asymmetric unit of BorB crystal contains A-chain (left) and B-chain (right) monomers. Light and dark blue regions identify the “core” while the orange and yellow regions identify the “lid”. Serine 98 is highlighted in red for context.



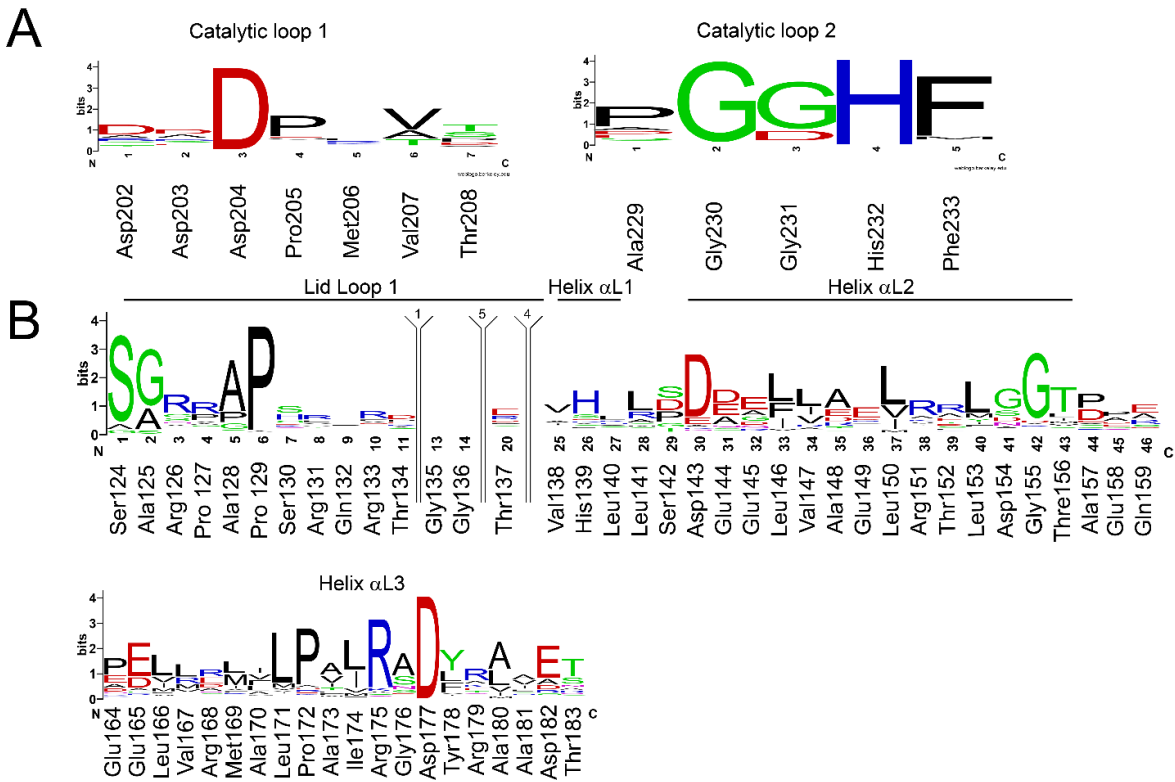
Supporting Figure 4-14. BorB A and B chains colored by B-factor. Greater B factors are depicted with warmer colors. Catalytic Ser98 is in purple.



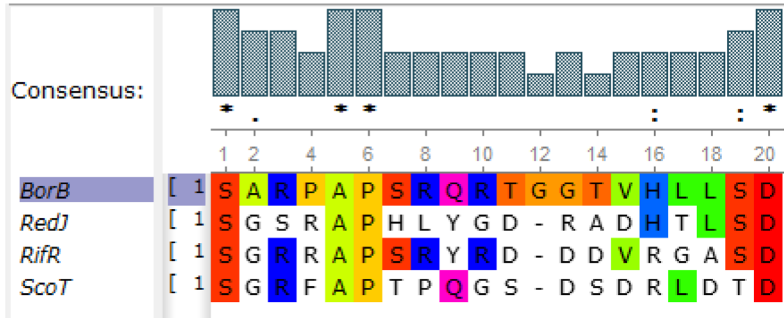
Supporting Figure 4-15. Array of BorB B chain and homologs co-crystallized with substrates and inhibitors. Active site serines colored in red.



Supporting Figure 4-16. BorB A chain with directionality of substrates indicated. S98 indicated for context.



Supporting Figure 4-17. A) Structure of BorB Catalytic Loop 1 and 2 with weblogo depicting residue conservation. B) Structure of BorB lid loop and helices with weblogo. (278)



Supporting Figure 4-18. Alignment of Lid Loop 1 between BorB, RedJ, RifR, and ScoT.

4.7. Acknowledgements

Marian-Joy Baluyot contributed significantly to cloning and kinetic experiments. Julie Lake and Hendrik Pütz provided general assistance. Jose Henrique Pereira performed the crystal screening, data collection, analysis and model generation for BorB. Structural experiments were conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Crystallography experiments were performed at beamline 8.2.2 of the ALS, a DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231, and were supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. Daniel Rosenburg and Michal Hammel performed SEC-SAX experiments at the SIBYLS beamline at the ALS. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. Jacquelyn Blake-Hedges for provided purified SFP. pRK793 was a gift from D. Waugh. We thank College of Chemistry's NMR facility for resources provided and the staff for their assistance. Instruments in CoC-NMR are supported in part by NIH S10OD024998.

4.8. Funding

This work was performed at the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. This research was also conducted as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies and Vehicle Technologies Offices.

Chapter 5. Engineering a chimeric borrelidin assembly line for methyl ketone production

5.1. Abstract

Developments in the last several decades of permitted the recombination of polyketide assembly lines for custom molecular production. A greater understanding of structural domain boundaries has been critical for engineering novel interactions without destabilizing pre-existing folds. In this chapter, we use some of these tools to combine modules from the borrelidin and lipomycin pathways to create methyl ketones. Using a combination of protein and small molecule mass spectrometry, we follow protein-bound intermediates through the biosynthetic pathway to the final production of two novel methyl branched methyl ketones.

5.2. Introduction

The iterative module BorM5 is capable of producing multi-methylated fatty acyl-ACPs (Chapter 3). (254) To produce free fatty acids, we attempted to use the Type II TE, BorB (Chapter 4). However, we discovered that BorB has a marked preference for short-chain acyl groups. Additionally, as discussed in Chapter 1, fatty acids do not make suitable biofuels due to their high melting temperatures, hygroscopicity, and viscosity. Ketones are an attractive alternative due to their favorable cetane numbers (2-undecanone has a CN of 56.6, while the minimum CN for diesel fuel is 40) and melting points.(38) Additionally, ketones often have pleasant smells and are used in cosmetics, fragrance, and food additives industries. (267) A fatty acid degradative approach, relying on \square oxidation, has been applied to produce methyl ketones (MKs) in the bacteria *E. coli*, (38) *Pseudomonas putida*, (41) and the *Ralstonia eutropha*, (268) and the fungus *Yarrowia lipolytica*. (269)

We therefore sought loading- and offloading partners for BorM5 to form a chimeric assembly line for methyl branched methyl ketones (MBMKs) (Figure 1-7). To engineer a loading module for BorM5, we chose the promiscuous lipomycin loading domain, which can provide a variety of short-chain substrates, including isobutyrate and isovalerate. (53) For offloading, we investigated the downstream module, BorM8. (175) BorM8 is the terminal module of the borrelidin synthase. It uses malonyl-CoA to extend, reduces to the β -keto moiety to a hydroxyl group, and contains a terminal TE for cyclization and release (Figure 3.1).

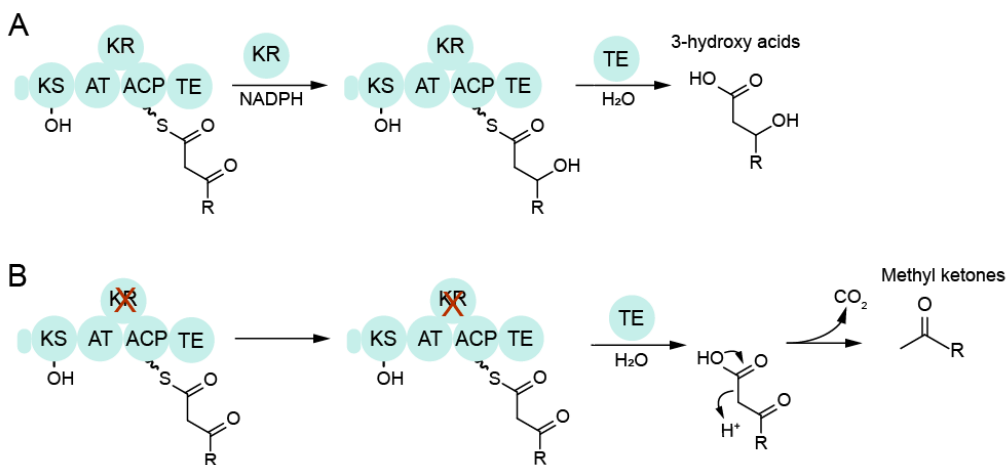


Figure 5-2. Methyl ketone production via KR knockout. **A)** Reduction of the β -ketone by the KR leads to β -hydroxy acids. **B)** KR inactivation precludes β -ketone reduction, leading to unstable β -keto carboxylic acids which decarboxylate non-enzymatically.

To produce ketones via polyketide biosynthesis, we imagined inactivating the ketoreductase (KR) domain). Polyketide biosynthesis proceeds through a β -keto intermediate, which the ketoreductase (KR) domain uses NADPH to reduce this moiety to an alcohol. The BorM8-TE natively releases a β -hydroxy acid (Figure 5-2, panel A). However, inactivation of the KR will yield a β -keto-acid after TE-mediated cleavage (Figure 5-2, panel B). β -keto acids decarboxylate non-enzymatically to afford ketones. (270) As a proof of principle, KR inactivation has been applied in engineered PKSs for the production of short-chain MKs in *Streptomyces albus*. (43)

5.3. Results

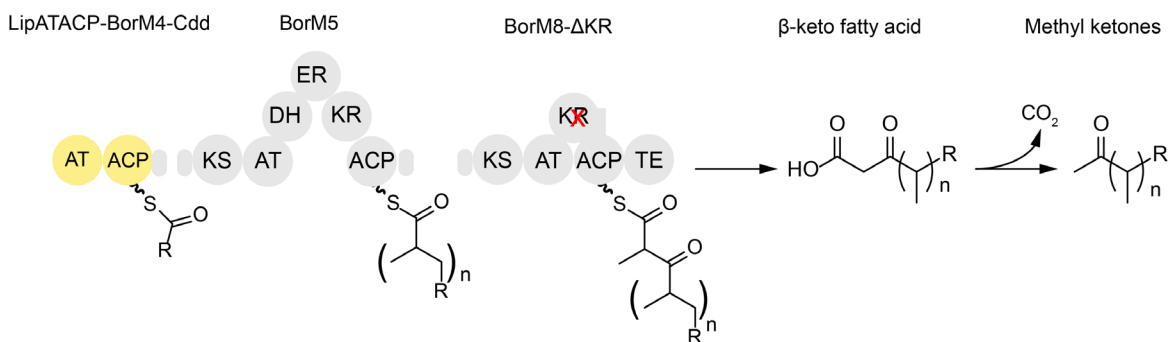


Figure 5-1: Engineered borrelidin pathway to produce methyl ketones.

5.3.1. Engineering a loading module

The lipomycin loading AT-ACP is natively part of the combined loading-extension module, LipM1. (53) The LipAT-ACP was therefore truncated after the conserved ACP domain. Rather than fuse LipAT-ACP to BorM5, for fear of disrupting protein folding and catalysis, we opted to engineer a *trans* interaction between the two partners. Polyketide synthase modules interact in *trans* via small C- and N-terminal linker domains, termed docking domains.(119) Fusion of compatible docking domains can promote interaction between non-native partners. (121, 122) In its assembly line, BorM5 is flanked by BorM4 at its N-terminus and BorM8 at its C-terminus (Figure 3.1). We therefore fused the BorM4 C-terminal docking domain (CDD) to the LipAT-ACP to form LipAT-ACP-BorM4-CDD.

To test if LipAT-ACP-BorM4-CDD was active and capable of transferring intermediates to BorM5 we reacted purified LipAT-ACP-BorM4-CDD *in vitro* and investigated acyl intermediates by LC-MS/MS-based phosphopantetheine ejection assay, which allows identification of ACP-bound intermediates. (202) Incubation of purified LipAT-ACP-BorM4-CDD with isobutyryl-CoA lead to the detection of isobutyryl-LipACP, indicating that the truncation did not disrupt native activity (Figure 5-3). Incubation of LipAT-ACP-BorM4-CDD with equimolar BorM5 and the substrates required for elongation (methylmalonyl-CoA and NADPH) lead to the detection of extension intermediates on BorM5-ACP (Figure 5-4, column B). Up to three extensions were observed, similar to extension of isobutyryl-SNAC (Figure 5-4, column A).

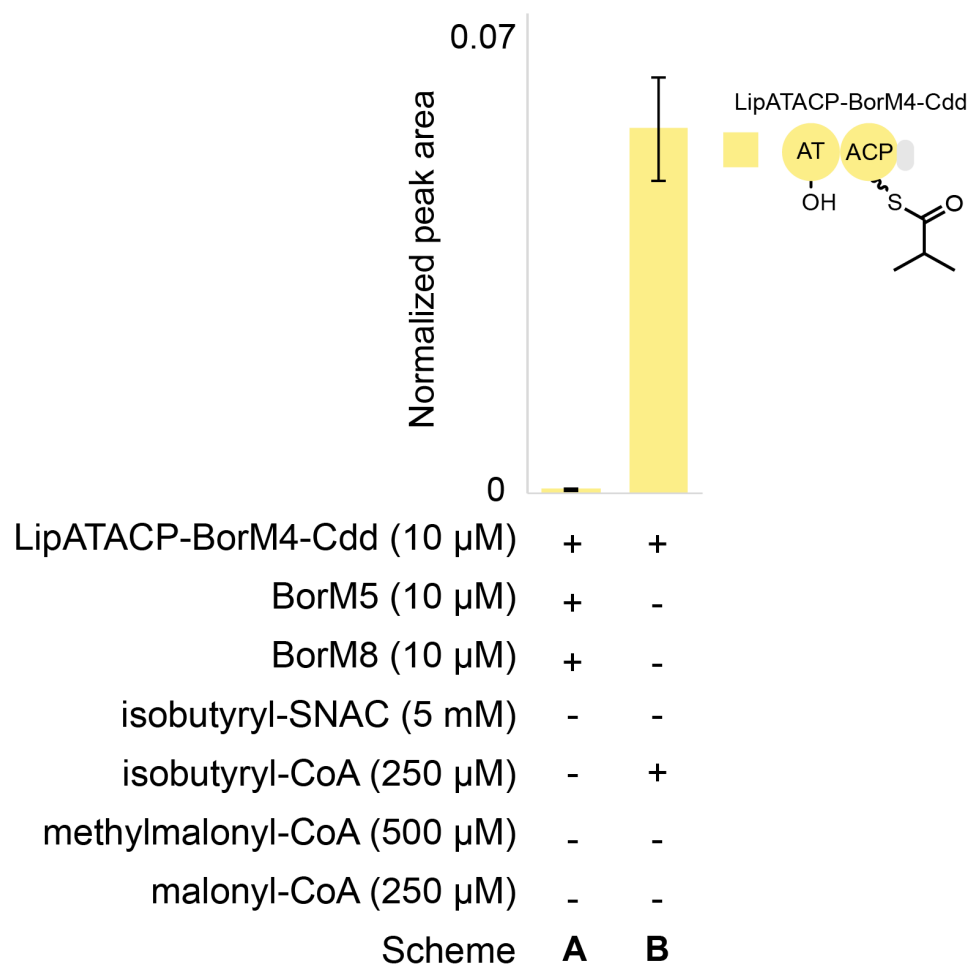


Figure 5-3. Self-acylation of purified LipAT-ACP-BorM4Cdd by LC-MS/MS. **A**) No substrates. **B**) Isobutyryl-CoA. All reactions contain 1 mM NADPH. Error bars are std. deviation of n=3 in vitro reactions.

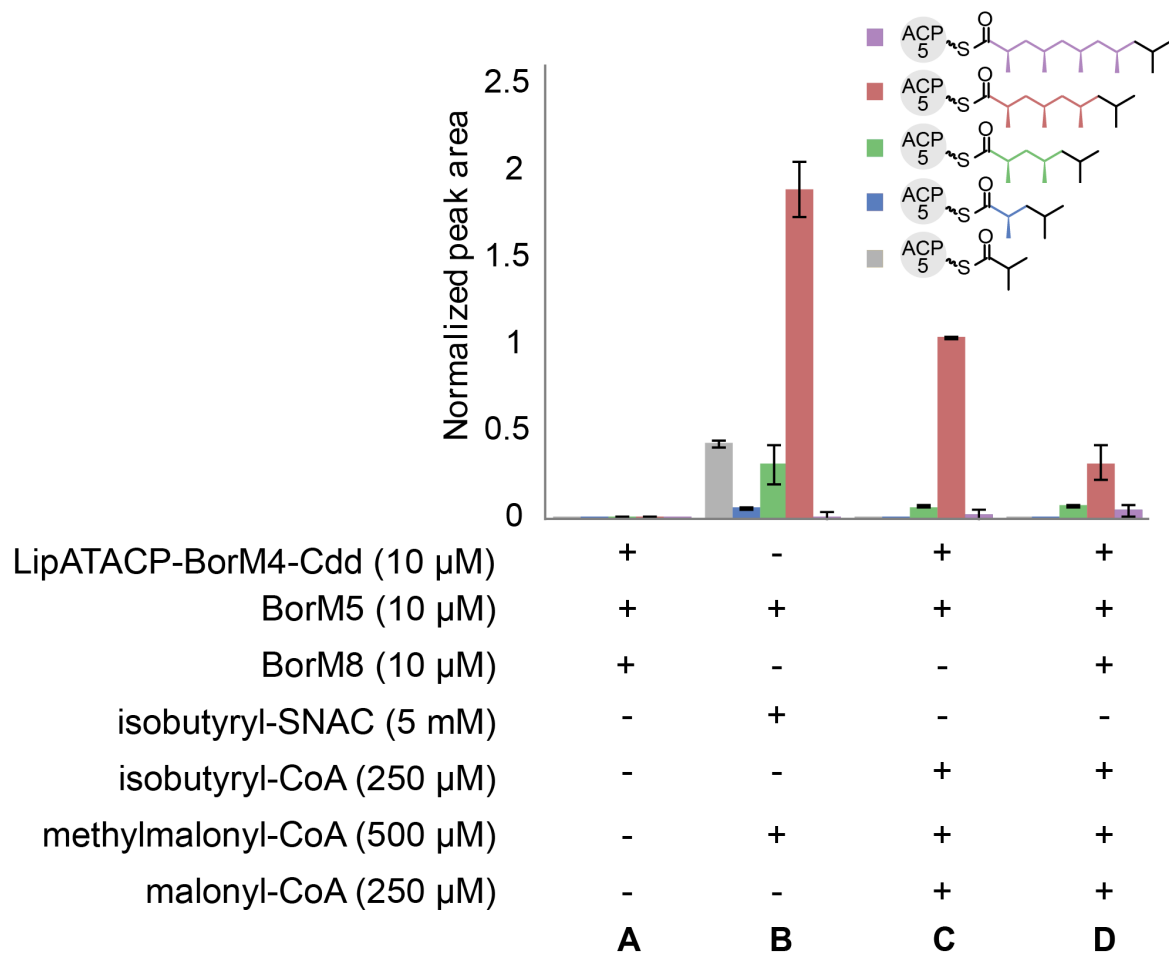


Figure 5-4. Production of fatty-acyl-ACP by BorM5 when provided with **A**), no substrates, **B**), extension substrates and isobutyryl-SNAC, **C**) isobutyryl-CoA, LipAT-ACP-BorM4-Cdd and extension substrates, or **D**) isobutyryl-CoA, LipAT-ACP-BorM4-Cdd and extension substrates and BorM8. All reactions contain 1 mM NADPH. Error bars are std. deviation of n=3 in vitro reactions.

5.3.2. Engineering an offloading module

BorM8, consisting of a KS-AT-KR-ACP-TE, natively elongates with malonyl-CoA and reduces the β -ketone to an alcohol via the KR domain (Figure 1). Thus, inactivation of the KR should lead to the production of a β -keto acid. However, we first investigated whether the native downstream module, BorM8 was capable of accepting and elongating intermediates from BorM5. BorM8 is a potential “gatekeeper” and may counter the inherent promiscuity of BorM5 with strict substrate specificity for the correct elongation intermediate. The KSs following the iterative AurA and NorA were highly specific for the correct chain length intermediates. (113, 114) To permit observation of intermediates, BorM8 was truncated after the conserved ACP domain to remove the TE, thereby limiting its ability to hydrolyze intermediates. BorM8- Δ TE-ACP elongation intermediates were observed when purified BorM8- Δ TE was incubated with isobutyryl-SNAC indicating functional protein (Supporting Figure 5-2).

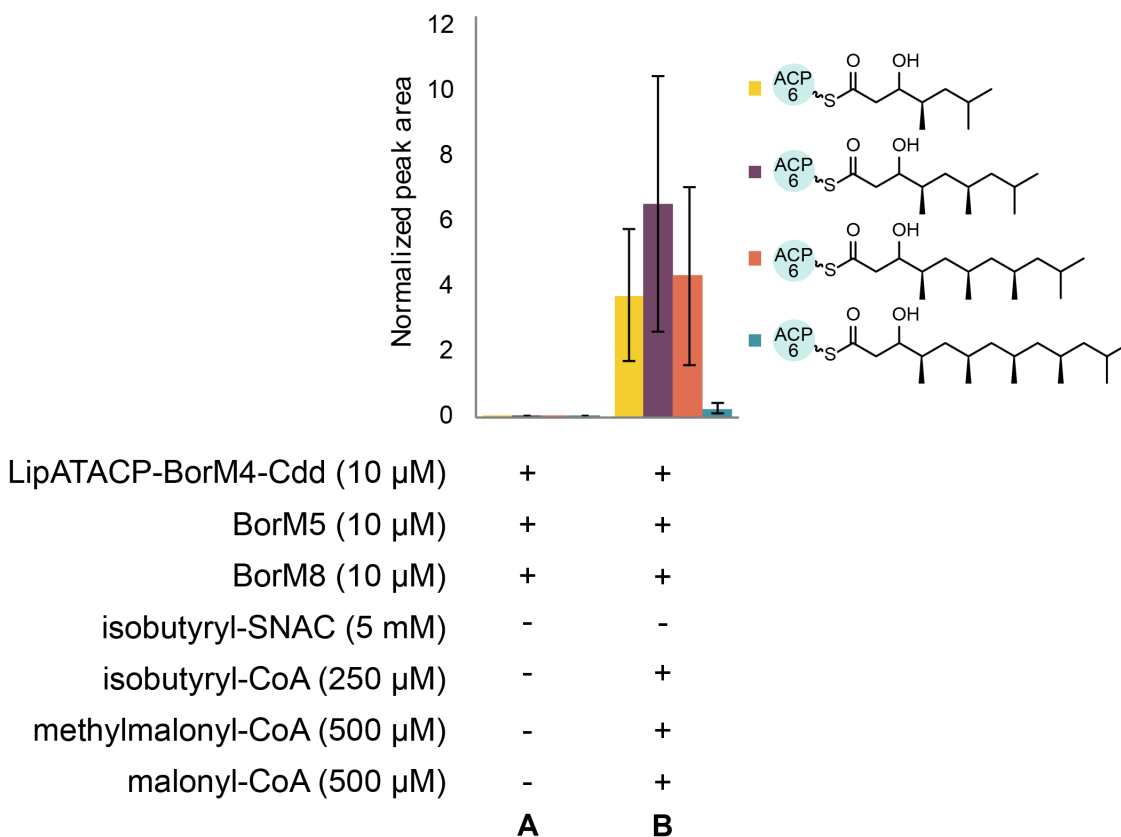


Figure 5-5. Production of fatty-acyl-ACP by BorM8 when provided with **A**), no substrates, or **B**) the full chimeric assembly line. All reactions contain 1 mM NADPH. Error bars are std. deviation of n=3 in vitro reactions.

When BorM8- Δ TE was incubated with LipAT-ACP-BorM4-CDD, BorM5, and the substrates required for elongation, we observed multiple intermediates on BorM8-ACP (Figure 5-5, column B). The masses of these intermediates indicated between 1-3 iterations by BorM5 and one extension by BorM8- Δ TE, confirming successful transfer between BorM5 and BorM8 as well as elongation by BorM8. Additionally, we observed the accumulation of 3-hydroxy-4-methylpentanoic-ACP, the product of BorM8 elongation of isobutyrate (Supporting Figure 5-3, column E). Incubation of BorM8- Δ TE and isobutyryl-CoA with or without LipAT-ACP-BorM4-CDD did not lead to appreciable isobutyryl-ACP8 (Supporting Figure 5-, column B and C). However, addition of BorM5 (but no elongation substrates) lead to an accumulation of 3-hydroxy-4-methylpentanoyl-ACP8 (Supporting Figure 5-3, column D). One possible and established mechanism for this is module “skipping” where the BorM5-ACP shuttles intermediates to BorM8 in the absence of elongation (Figure 5-6). (203)

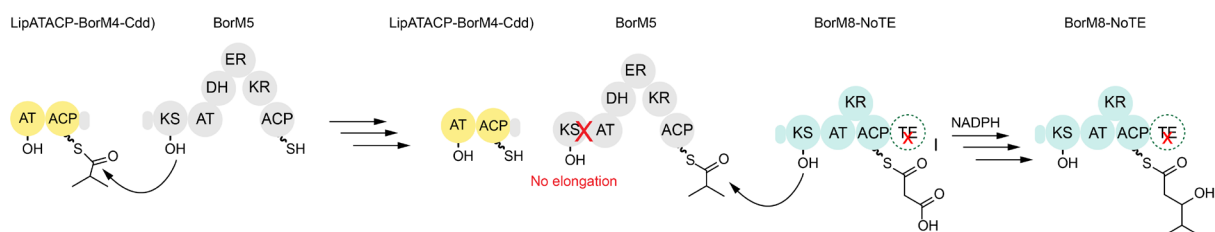


Figure 5-6. Potential skipping mechanism leading to the accumulation of 3-hydroxy-4-methylpentanoyl-BorM8-ACP.

5.3.3. Production and detection of methyl ketones in vitro

Having established transfer and elongation of intermediates from LipAT-ACP-BorM4-CDD to BorM8- Δ TE, we next generated a reduction-incompetent mutant of the TE-intact BorM8 by mutagenesis of catalytic tyrosine to phenylalanine (Y1246F). (140) Incubation of BorM8-Y1246F, but not BorM8 or BorM8- Δ TE, with LipAT-ACP-BorM4-CDD, BorM5, and elongation substrates lead to the GC-MS detection of several new peaks (Figure 5-7).

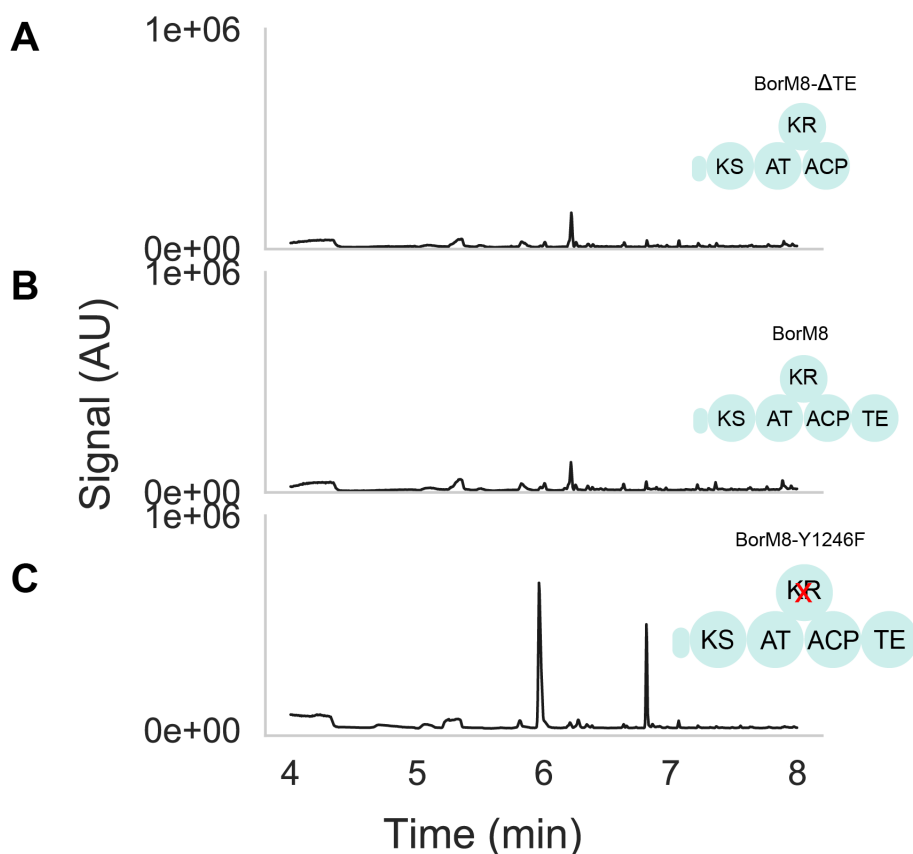
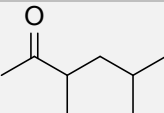
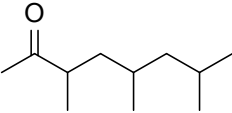
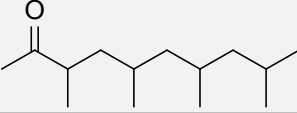
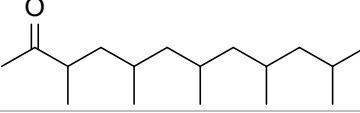


Figure 5-7. GC-MS total ion chromatograms of in vitro reactions with methylmalonyl-CoA, malonyl-CoA, isobutyryl-CoA, NADPH, LipAT-ACP-BorM4CDD, BorM5, and the BorM8 variant indicated. **A)** BorM8- Δ TE. **B)** WT BorM8. **C)** BorM8-Y1246F.

These peaks were analyzed for masses corresponding to the molecular ions and fragments expected for methyl ketones. The expected masses for MBMKs are displayed in Table 5-1. Extracted ion chromatograms for predicted fragments and molecular ions are shown in Figure 5-8. A slight peak 4.68 minutes contained-barely detectable potential ions **5** and **6** and a peak for 128 Da **1** (Figure 5-8). Peaks at 5.96 and 6.81 minutes contained strong signals for ions **5** and **6** as well as signals for 170 and 212 Da for MBMKs **2** and **3** respectively (Figure 5-8). The spectra for these peaks (Figure 5-9) were compared to the NIST Spectral database and matched with generic α -methylated methyl ketones. The 254 mass, corresponding to **4**, was not observed.

Table 5-1. Expected products.

#	Name	Structure	Mass (Da)	Formula
1	3,5,7-dimethylhexan-2-one		128.12	C ₈ H ₁₆ O
2	3,5,7-trimethyloctan-2-one		170.17	C ₁₁ H ₂₂ O
3	3,5,7,9-tetramethyldecane-2-one		212.21	C ₁₄ H ₂₈ O
4	3,5,7,9,11-pentamethyldodecan-2-one		254.26	C ₁₇ H ₃₄ O

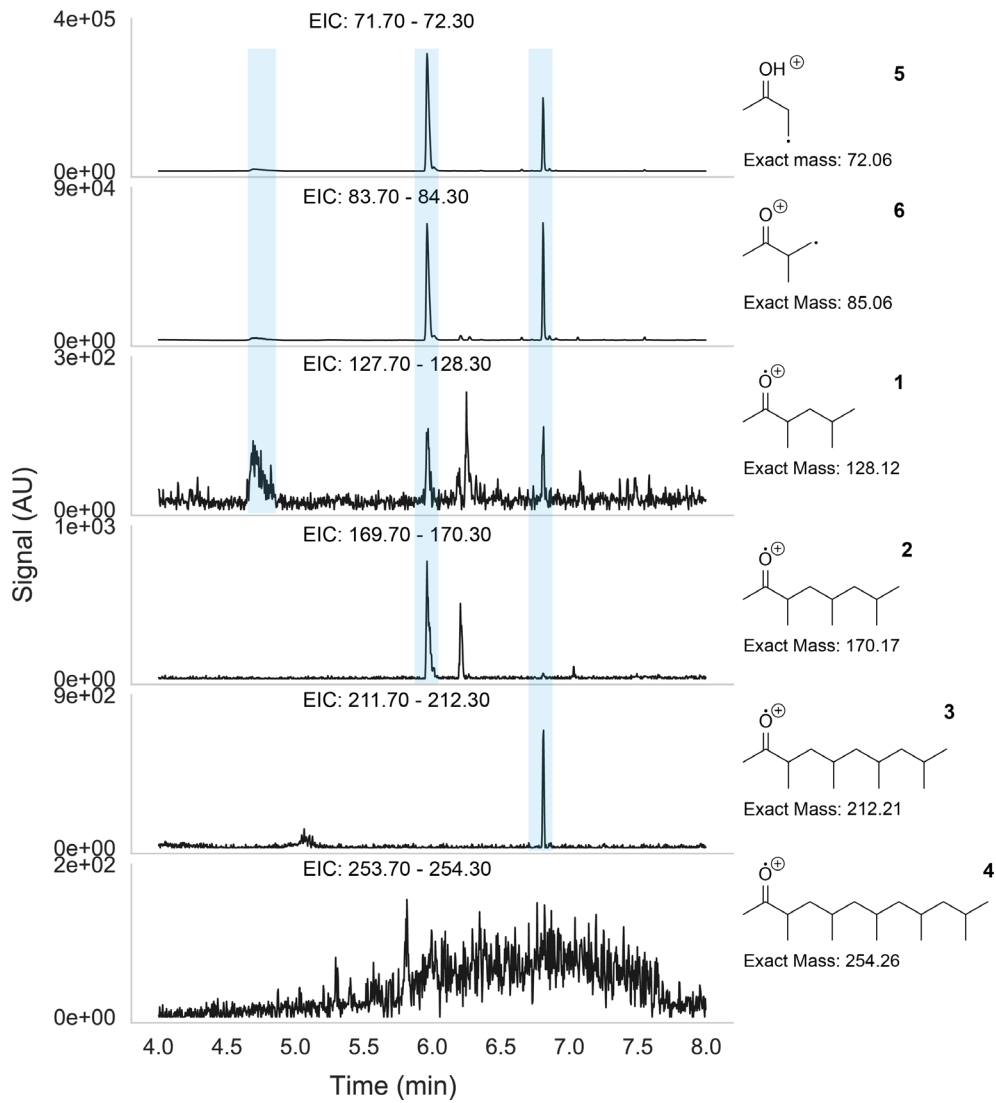


Figure 5-8. Extracted ion chromatograms for the BorM8-Y1246F reaction. Masses shown above each trace. Putative fragments and molecular ions shown on the right. Blue boxes highlight peaks of interest.

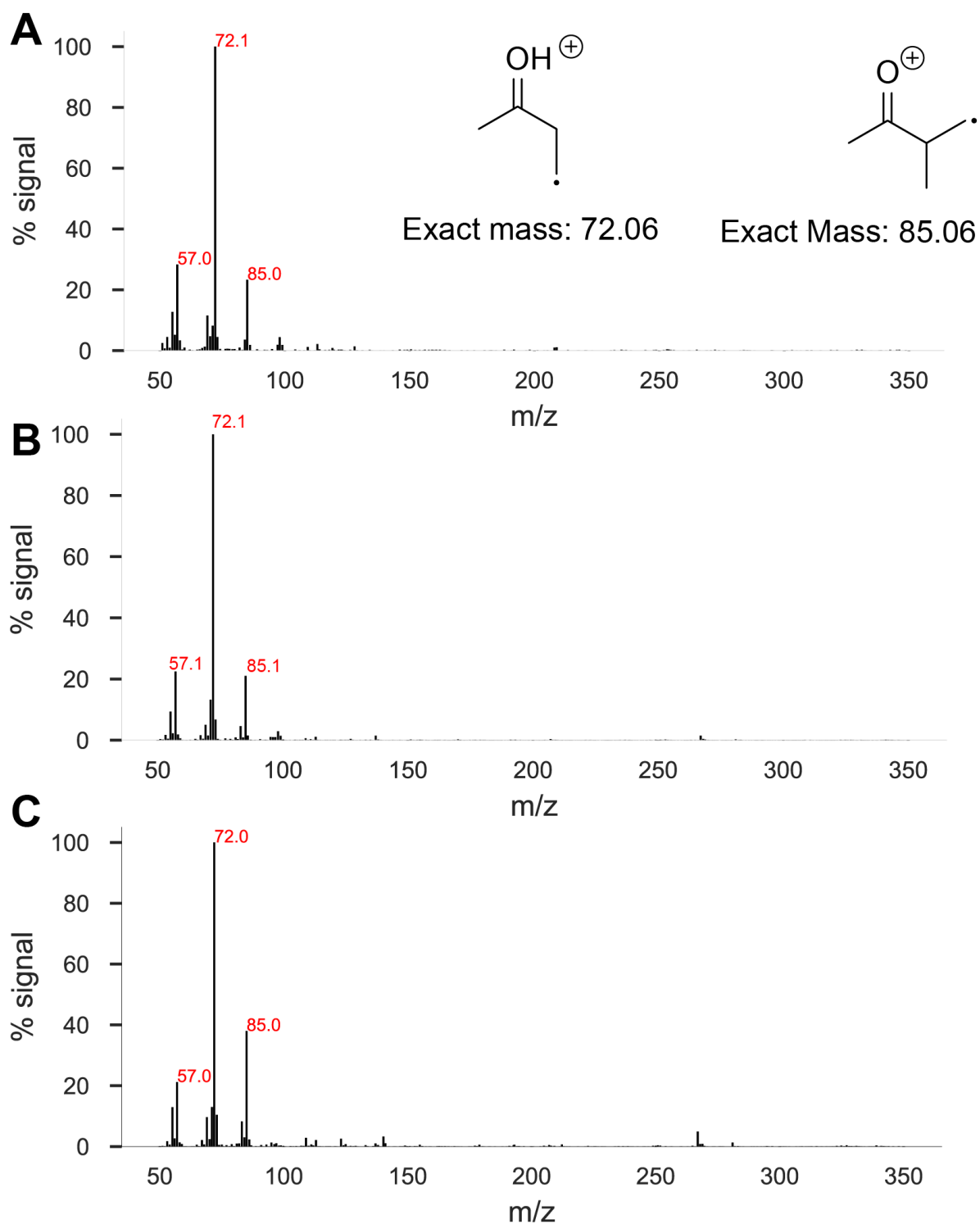


Figure 5-9. Spectra of peaks of interest. **A.** Peak at 4.68 min. Several putative ions are shown. **B)** Peak at 5.96 min. **C)** Peak at 6.81 min.

5.4. Discussion

Here we have applied a combination of protein- and GC-MS to analyze intermediates and products of a chimeric assembly line. The phosphopantetheine ejection assay was used to follow intermediates through thiotemplated polyketide biosynthesis. Only reactions containing a non-reducing terminal module produced novel MBMKs. The GC-MS spectra for these compounds are consistent with the production of primarily C11 and C14 MBMKs 3,5,7-trimethyloctan-2-one (**2**) and 3,5,7,9-tetramethyldecan-2-one (**3**).

5.5. Methods

5.5.1. Reagents

All reagents were purchased through Sigma unless otherwise noted.

5.5.2. Acyl-SNAC synthesis

Isobutyryl-SNAC was previously synthesized. (254)

5.5.3. Acyl-CoAs

All acyl-CoAs were purchased from Coala Biosciences.

5.5.4. Enzyme domain boundaries

Domains were defined using the MiBIG database. (252)

5.5.5. Plasmid Construction

PCRs were designed using the DeviceEditor GUI and J5 assembly. (210, 213) All PCRs were accomplished using Q5 polymerase (NEB) and digested with DpnI before assembly. Fragments were purified or gel extracted as needed (Qiagen). PCR fragments were assembled using Gibson mix (NEB) or T4 ligase (NEB) as indicated for each plasmid. (214) Assemblies were transformed into DH10b (UC Berkeley MacroLab) unless otherwise noted. Sanger sequencing was performed by Sequetech (Mountain View, CA). For plasmids made in this study, the sequences, primer sequences, and physical samples are available via the Joint BioEnergy Institute Registry (Supporting Table 5-1).

Supporting Table 5-1. Plasmids.

Name	Source	Registry link
pARH130 (pBbS2k-6xHis-MBP-BorM5)	Curran, 2018.(254)	https://registry.jbei.org/entry/33363
pARH134 (pBbS2k-6xHis-MBP-BorM8)	This study	https://registry.jbei.org/entry/45093
pARH135 (pBbS2k-6xHis-MBP-BorM8-NoTE)	This study	https://registry.jbei.org/entry/45094
pSC054 (pBbS2k-6xHis-MBP-BorM8-Y1246F)	This study	https://registry.jbei.org/entry/104541
pHP02 ((pBbS2k-6xHis-MBP-LipAT-ACP-BorM4-CDD)	This study	https://registry.jbei.org/entry/100107
pRK793 (TEV protease)	David Waugh(211)	https://www.addgene.org/8827/
pET-SFP	Jacquelyn Blake- Hedges	https://registry.jbei.org/entry/58169

pARH130 (pBbS2k-6xHis-MBP-borM5)

pARH130 was previously constructed. (254)

pARH134 (pBbS2k-6xHis-MBP-borM8) The borM8 open reading frame was codon optimized for *E. coli* and synthesized (GeneWiz) and provided in pUC57. pUC57-BorM8 was digested with NdeI/XhoI and ligated into a similarly-digested pBbS2K backbone. (212)

pARH135 (pBbS2k-6xHis-MBP-borM8-NoTE)

A silent BamHI site was installed immediately after the conserved ACP during synthesis of the borM8 gene. The pUC57-borM8 construct was digested with NdeI/BamHI and ligated into a similarly-digested pBbS2K backbone. (212)

pSC054 (pBbS2k-6xHis-MBP-borM8-Y1246F)

The tyrosine-to-phenylalanine point mutant was installed into pARH134 via two overlapping primers in a single elongation. (271)

pHP02 (pBbS2k-6xHis-MBP-lipAT-ACPborM4-CDD)

The DNA coding for the N-terminally truncated lipomycin module 1 loading didomain was obtained from the JBEI registry. (272) Transcription starting site 4 was chosen as N-terminal truncation site and the end of the conserved AT was chosen as the C-terminus. The C-terminal docking domain of borM4 was codon optimized for *E. coli* and synthesized (GeneWiz). Fragments were PCR amplified and assembled in a NdeI/XhoI digested pBbS2K backbone. (212)

5.5.6. Protein purification

BorM5 and BorM8 variants

Bl-21 (DE3) cells containing expression plasmids were cultured overnight at 37°C, 200 rpm in 50 mL Terrific Broth (TB), 50 µg*mL⁻¹ kanamycin before being inoculated 1% v/v into 2 L TB media, 50 µg*mL⁻¹ kanamycin. (273) Cells were split into four 2 L baffled shake flasks and incubated at 37°C, 200 rpm until they reached OD₆₀₀ ≥ .2 AU. Protein expression was induced with 50 ng*mL⁻¹ anhydrotetracycline and cells were incubated at 18°C, 200 rpm for 20 hours. Cells were centrifuged (6000 × g, 10 min, 4°C), the supernatant was discarded, and the cells were suspended in 50 mL lysis buffer (50 mM NaPO₄²⁻, pH 7.5, 150 mM NaCl, 20 mM imidazole). Cells were stirred at 200 rpm on ice until clumps were no longer visible (15-20 minutes). Cells were lysed by 5-10 passes through an Avestin Emulsiflex C3 homogenizer (15,000 bar) and centrifuged (40,000 × g, 30 min, 4°C) to pellet the insoluble fraction.

NiNTA purification was done using 5 mL GE HisTrap column on an AktA FPLC. The column was equilibrated in 10 CV lysis buffer, 10 CV elution buffer (50 mM NaPO₄²⁻,

pH 7.5, 150 mM NaCl, 500 mM imidazole), and 10 CV lysis buffer. The soluble fraction of the lysate was loaded onto the column via the sample pump at 2 mL*min⁻¹ and the FT was collected. The column was washed with 10 CV lysis buffer. The bound protein was eluted at 2.5 mL*min⁻¹ over 100 mL. The desired fractions were identified by SDS-PAGE.

~20 mL protein from NiNTA was dialyzed in dialysis tubing (Thermo) in 1 L AEC buffer A (50 mM NaPO₄²⁻, pH 7.5, 50 mM NaCl, 1 mM DTT) for 1 hour at 4°C. This was repeated. AEC purification was done using a 5 mL HiTrap Q anion column on an Akta FPLC. The column was equilibrated in 10 CV AEC buffer A (50 mM NaPO₄²⁻, pH 7.5, 50 mM NaCl, 1 mM DTT) 10 CV AEC buffer B (50 mM NaPO₄²⁻, pH 7.5, 1 M NaCl, 1 mM DTT). The dialyzed protein sample was loaded via the sample pump at 5 mL*min⁻¹. The bound protein was washed with 10CV buffer A. The bound protein was eluted in a gradient (0-100% buffer B over 60 minutes at 1 mL*min⁻¹). The desired fractions were identified by SDS-PAGE.

The protein concentration in the pooled AEC fractions was quantified by UV absorbance and checked for purity via SDS-PAGE (Supporting Figure 5-1, panels B-E). The sample was concentrated, buffer exchanged by spin filtration into 50 mM NaPO₄²⁻, pH 7.5, 8% glycerol, 1 mM DTT, flash frozen in liquid nitrogen, and stored at -80°C.

LipAT-ACP-BorM4-CDD

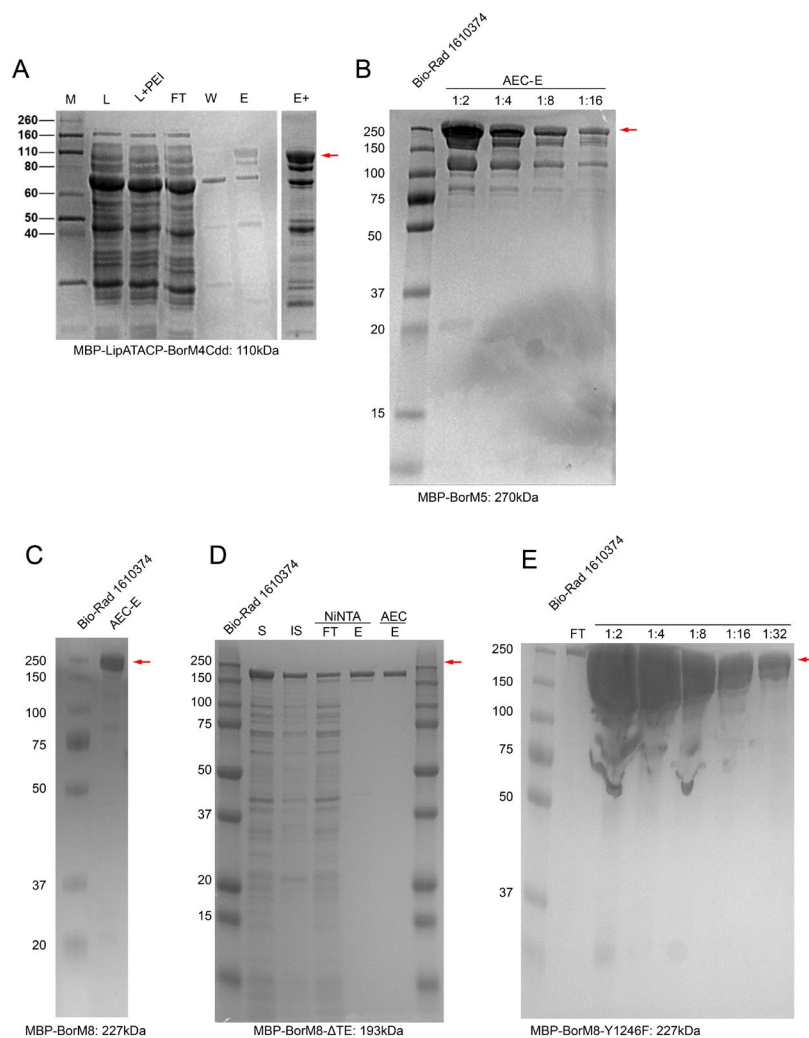
LipAT-ACP-BorM4-CDD was purified only by NiNTA chromatography as described above (Supporting Figure 5-, panel A).

TEV protease

The purification for the TEV protease followed the original published protocol. (211)

SFP

SFP was kindly provided by Jacquelyn Blake-Hedges.



Supporting Figure 5-1. Purified proteins used in this study. **A**) MBP-LipAT-ACP-BorM4Cdd Ni-NTA purification. M, marker; L, lysate; L+PEI, lysate with polyethylenimine; W, wash; E, elute, **B**) dilution series of MBP-BorM5 anion exchange elutions (AEC-E), **C**) MBP-BorM8 anion exchange eluent, **D**) MBP-BorM8 Δ TE NiNTA and AEC purification. S, soluble cell material; IS, insoluble material **E**) MBP-BorM8-Y1246F dilution series.

5.5.7. In vitro assays

Reactions were performed in 100 mM Na₂NaPO₄²⁻, pH 7.6, 10 mM MgCl₂, 1 mM TCEP. Biosynthetic enzymes were added, as indicated, to a final concentration of 10 μ M. Prior to substrate addition, TEV-cleavage and phosphopantetheinylation were performed by

the addition of 5 μ M purified TEV, 1 μ M purified SFP, and 100 μ M CoA-SH. This reaction proceeded for 1 hour at 25°C. Substrates were added at the following concentrations: NADPH, 1 mM; isobutyryl-CoA, 0.5 mM, malonyl-CoA, 0.5 mM; methylmalonyl-CoA, 1 mM. Reactions proceeded at 25°C for 2 hours noted before freezing, unless otherwise noted.

5.5.8. Protein mass spectrometry

Sample preparation, LC-MS/MS sample run, data acquisition, and processing were performed as described in Chapter 3. The raw data is available on the public mass spectrometry repository Panorama, at the URL: <https://panoramaweb.org/project/JBEI/begin.view?>. (219) Links for individual figures are shown in Supporting Table 5-2.

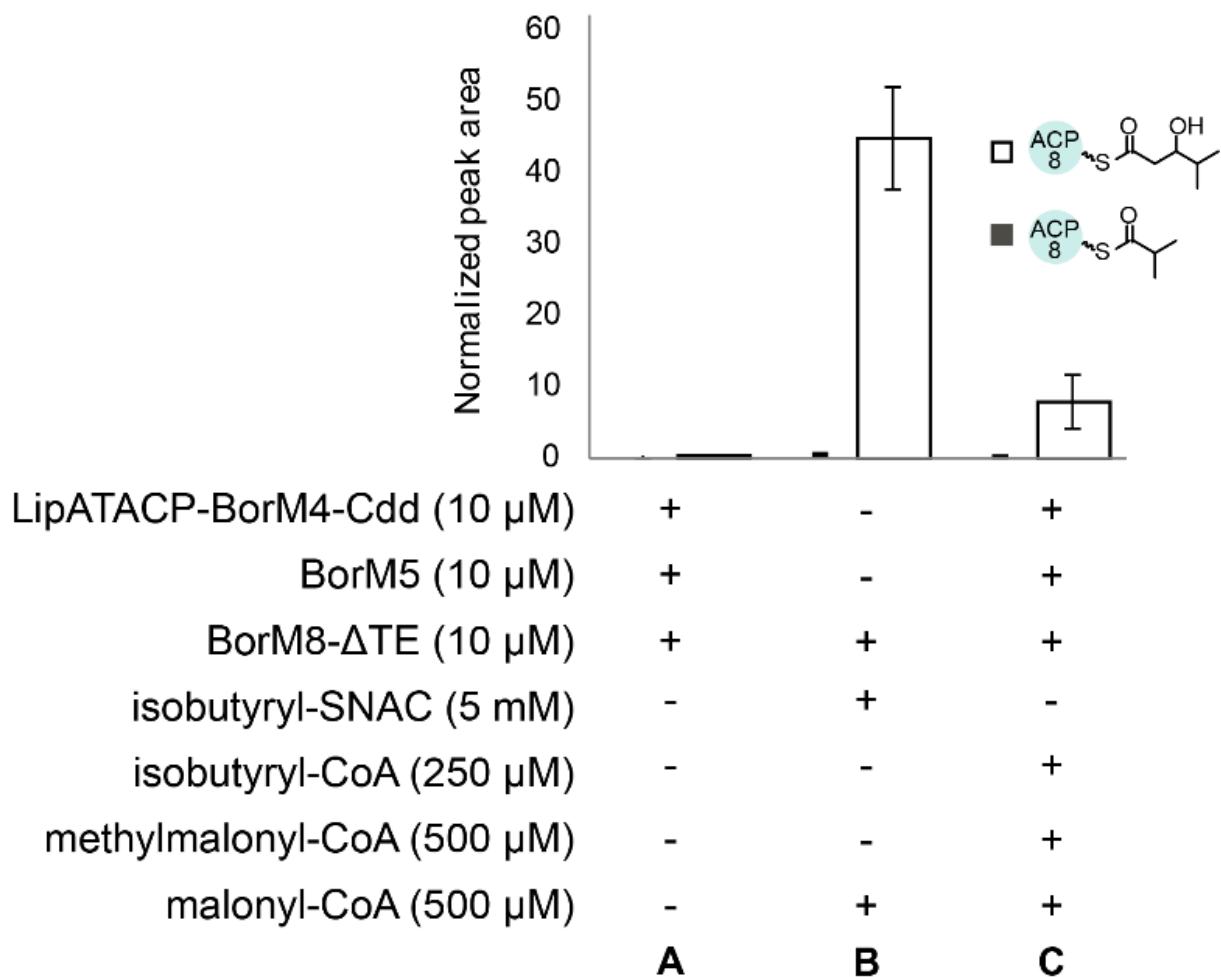
Supporting Table 5-2. Raw data links.

Figure #	Panorama Link
1-3	20190130 hp02 borm5 borm8 isobutyryl_2019-11-01_18-05-53.sky.zip
1-4	20190130 hp02 borm5 borm8 isobutyryl_2019-11-01_18-05-53.sky.zip
1-5	20190130 hp02 borm5 borm8 isobutyryl_2019-11-01_18-05-53.sky.zip
1-6	20190130 hp02 borm5 borm8 isobutyryl_2019-11-01_18-05-53.sky.zip
1-7	20190204 hp02 borm5 borm8 isobutyryl_2019-11-01_18-06-43.sky.zip

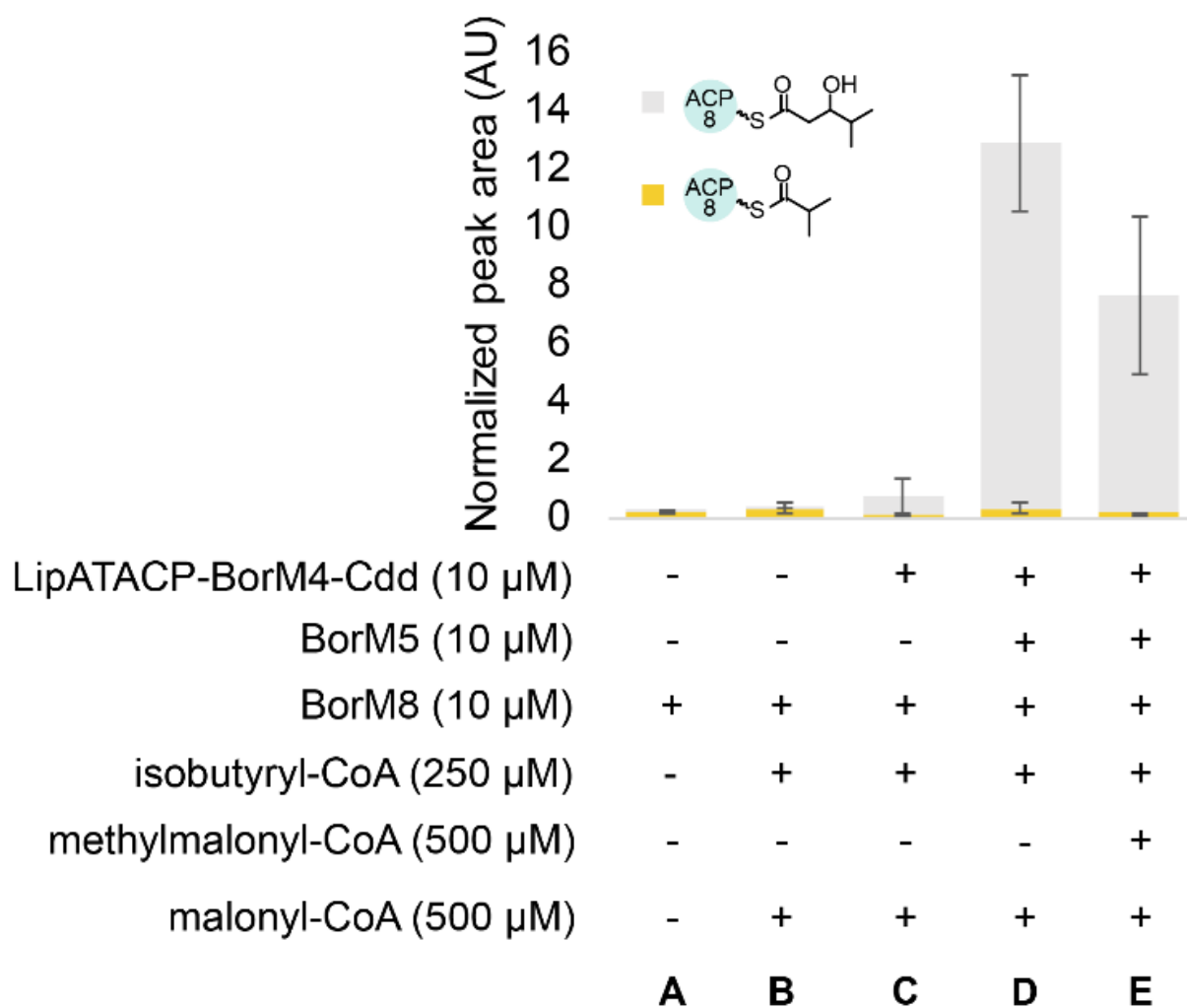
5.5.9. GC-MS

GC-MS for methyl ketones was performed via headspace analysis. 1 mL enzymatic reactions were set up in gas-tight vials with Mininert valves (Supelco). 100 μ L headspace was directly injected onto an Agilent 7890A GC coupled to a 5975C MSD equipped with a HP-5 mS 5% phenyl methyl silox column. The GC was run in splitless mode. Inlet and transfer line temperature were held constant at 250°C and 280°C, respectively. The oven was held at 50°C for 3 minutes and then increased to 300°C at a rate of 50°C*min⁻¹. Oven temperature was held at 300°C for 5 minutes. A 2-minute solvent delay was used. The MS was run in scan mode from 50.0 to 350.0 Da at 1212 EM volts with a gain factor of 1.82. Data was analyzed via Chemstation Enhanced Data Analysis program. Structural predictions were performed using the NIST Mass Spectral Search Program.

5.6. Supporting Figures



Supporting Figure 5-2. Elongation of isobutyrate by BorM8. **A**) no substrates, **B**) isobutyryl-SNAC, malonyl-CoA, **C**) full assembly line. All reactions contain 1 mM NADPH. Error bars are std. deviation of n=3 in vitro reactions.



Supporting Figure 5-3. LC-MS/MS of BorM8-ACP bound isobutyrate and 3-hydroxy-4-methylpentanoate with, **A**) no substrates, **B**), BorM8, isobutyryl-CoA, and malonyl-CoA, **C**) B with addition of LipAT-ACP-BorM4-Cdd, **D**) C with the addition of BorM5, and **E**) D with the addition of methylmalonyl-CoA. All reactions contain 1 mM NADPH. Error bars are std. deviation of n=3 in vitro reactions.

5.7. Acknowledgements

Hendrik Pütz generated significant experimental findings and contributed to the intellectual development of this work. Marian-Joy Baluyot and Julie Lake Provided essential experimental assistance. The protein mass spectrometry was performed with the assistance of Christopher Petzold and Yan Chen at the Joint BioEnergy Institute. Jim Kirby and Andria Rodriguez provided GC-MS guidance. Jacquelyn Blake-Hedges provided purified SFP. pRK793 was a gift from David Waugh.

5.8. Funding

This work was performed at the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. This research was also conducted as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies and Vehicle Technologies Offices. Hendrik Pütz contributed to this work as part of the German Academic Exchange Service with resources from the German Federal Ministry of Education and Research.

Chapter 6. Conclusion

In this dissertation work we assembled a chimeric PKS assembly line to produce MBMKs. The biosynthetic chassis is the iterative/modular BorM5, of which little was known beforehand. We used the lipomycin loading AT-ACP to prime biosynthesis. Fusion of a compatible linker domain was necessary for these modules to interact productively. We used the downstream module BorM8 to accept and elongate intermediates and offload products. Deactivation of the KR domain yielded α -keto acids and their subsequent decarboxylation to methyl ketones. We additionally investigated the *trans*-acting thioesterase, BorB, as a chain-releasing mechanism. However, we discovered that BorB prefers short-chain substrates, limiting its utility for biodiesel molecules.

We demonstrate that a single iterative module can replace several co-linear modules. This may become a feasible strategy for engineering minimized assembly lines. One downside is that the iterative module performs the same modifications with each elongation, which is not always desirable. AT-swapping, as demonstrated here, permits the use of promiscuous ATs and can generate diverse intermediates. Additionally, many iterative modules, especially fungal PKSs, perform variable modifications on substrates using the same set of domains. (75) The “programming” of these reactions is at the forefront of understanding PKS biochemistry.

The downstream module plays a significant role in product formation. KS domains from canonical *cis*-AT PKSs are generally promiscuous, with the exception of several cases of stereocontrol. (135) This is a double-edged sword; if intermediate transfer to the downstream KS_{n+1} is faster than retro-transfer to the KS_n , then the intermediates will be offloaded before the desired chain length can be attained. The presence of multiple intermediates being passed to BorM8 may reflect the kinetic competition between the two modules. Interestingly, the modules that natively follow iterative modules are thought to be highly-specific “gatekeepers”. (114) Our findings that BorM8 accepts a wide range of short- and long-chain acyl substrate question the absolute specificity of this module. However, BorM8 may still act as a gatekeeper in its native setting, only activating the third elongation intermediate from BorM5.

This work was carried out entirely *in vitro* using purified proteins and substrates. We employed this strategy due, 1) the lack of certainty on identity of the target molecules due to the unpredictability of BorM5 and the inherently low background of *in vitro* reactions, 2) the prohibitive cost of chemical synthesis of analytical standards, 3) the degree of control afforded, and 4) the ease of protein mass spectrometry for enzyme-bound intermediates. Indeed, the phosphopantetheine ejection assay informed most experiments in the biosynthesis of MBMKs. This allowed us to probe the activity of each

enzyme in the assembly line rather than rely on product formation, which is dependent on all pathway enzymes.

Using the phosphopantetheine ejection assay, we discovered that BorM5 is promiscuous for both priming and elongation substrates and iterates a different number of times depending on the substrate, similar to other iterative PKSs reviewed in Chapter 2. We used the phosphopantetheine ejection assay to confirm transfer of intermediates between modules LipAT-ACP-BorM4Cdd:BorM5 and BorM5:BorM8- Δ TE. The phosphopantetheine ejection assay also identified several bottlenecks in the pathway that may limit yields. The first is that BorM5 is inhibited by the starting substrate. This is logical as the KS is the site of both initial priming and iterative retrotransfer. Similarly, fatty acid synthases are also substrate-inhibited. (206) The second area for improvement is the “module skipping” observed in Chapter 5. The priming substrate, isobutyrate, was transferred from LipAT-ACP-BorM4Cdd to BorM8 via BorM5 without elongation. This has been observed in several other PKSs. (203)

While we have demonstrated the production of two MBMKs in vitro, much remains to be done. The current molecular portfolio can be expanded by altering the PKS assembly line. Different loading domains can provide different priming substrates to BorM5, leading to a variable number of elongations and therefore chain branches. AT swapping can also lead to variable branching patterns, such as desmethylation, gem-dimethylation, or ethylation. Additional PKS domains could be used to form rings and branches.

Yet this molecular diversity cannot be realized without a means to make the molecules in bulk. Future efforts will be diverted towards developing a high-titer in vivo expression system. Many hosts have been used to produce polyketides, including *E. coli* and *S. cerevisiae*. (273, 274) Streptomycetes have also attracted much attention in this field. Streptomycetes natively produce natural products, with up to dozens of clusters in their genomes. (275) Compared to *E. coli*, they offer native post-translational modification of polyketide pathways and a diverse pool of cofactors and substrates. (276) Methylmalonyl-CoA, for example, is endogenously produced in streptomycetes but not *E. coli*. (273) *E. coli* also lacks the ability to produce isobutyryl-CoA. (29) The enzymes used in this study come from streptomycetes; borrelidin is produced by two streptomycetes, *S. parvulus* and *S. rocheii*, among other species. (175) Lipomycin is produced by *S. aureofaciens* Tü117. (277) Streptomycetes have also been used for the production of short-chain ketones, reaching over 1 g*L⁻¹ of ethyl ketones. (43)

The establishment of a system to both diversify and produce MBMKs is the first stage in investigating the reality of MBMKs as biodiesels. As this chemical space is largely unexplored, several rounds of testing and modifying may be required to find molecules

with desirable properties. This work may also improve our ability to predict fuel properties from structure alone.

In conclusion, this dissertation lays the groundwork for an iterative polyketide biosynthetic assembly line to produce methyl-branched methyl ketones. The modular nature of the assembly line permits the mixing and matching of components for tailored products. This work highlights the flexibility and power of biology, while chronicling some of the challenges to control it. Biology can be applied to replace existing petroleum-derived products, leading to reductions in carbon emissions and improvements in sustainable industry. Further, as demonstrated here, biology can produce new structures, replete with untested material properties. The petroleum revolution changed the world with cheap portable energy and durable goods. New-to-nature biomolecules may offer that same promise, waiting to be discovered.

Chapter 7. References

1. Dodson J, et al. (2014) Use of coal in the Bronze Age in China. *The Holocene* 24(5):525–530.
2. Smith AHV (1997) Provenance of Coals from Roman Sites in England and Wales. *Britannia* 28:297.
3. First American Oil Well - American Oil & Gas Historical Society Available at: <https://aoghs.org/petroleum-pioneers/american-oil-history/> [Accessed October 17, 2019].
4. History of gasoline - U.S. Energy Information Administration (EIA) Available at: <https://www.eia.gov/energyexplained/gasoline/history-of-gasoline.php> [Accessed October 17, 2019].
5. Feldman D (2008) Polymer History. *Designed Monomers and Polymers* 11(1):1–15.
6. United Nations [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)] (2014) (United Nations, Geneva, Switzerland), p 151.
7. The California Water System *The California Water System*. Available at: <https://water.ca.gov/Water-Basics/The-California-Water-System> [Accessed October 17, 2019].
8. Obama B (2017) The irreversible momentum of clean energy. *Science* 355(6321):126–129.
9. Horowitz CA (2016) Paris Agreement. *International Legal Materials* 55(4):740–755.
10. Bottino NR (1971) The composition of marine-oil triglycerides as determined by silver ion-thin-layer chromatography. *J Lipid Res* 12(1):24–30.
11. Monthly Biodiesel Production Report - Energy Information Administration Available at: <https://www.eia.gov/biofuels/biodiesel/production/> [Accessed October 17, 2019].
12. How much ethanol is in gasoline, and how does it affect fuel economy? - FAQ - U.S. Energy Information Administration (EIA) Available at: <https://www.eia.gov/tools/faqs/faq.php?id=27&t=10> [Accessed October 17, 2019].
13. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci U S A* 103(30):11206–11210.
14. Robertson GP, et al. (2017) Cellulosic biofuel contributions to a sustainable energy future: Choices and outcomes. *Science* 356(6345).
15. Chevron Global Marketing (2007) *Diesel Fuels Technical Review* (Chevron Corporation).
16. Knothe G (2005) Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Processing Technology* 86(10):1059–1070.

17. Ng T-B ed. (2011) *Soybean: Applications and Technology* (BoD – Books on Demand).
18. Biodiesel K, content M, blends. B ABSTRACT. Biodiesel has the characteristic of absorbing more moisture than petroleum diesel. High moisture content in.
19. Dunn RO, Knothe G (2001) Alternative Diesel Fuels from Vegetable Oils and Animal Fats. *J Oleo Sci* 50(5):415–426.
20. Knothe G, Dunn RO (2009) A comprehensive evaluation of the melting points of fatty acids and esters determined by differential scanning calorimetry. *J Am Oil Chem Soc* 86(9):843–856.
21. Lander ES, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.
22. Kelley NJ, et al. (2014) Engineering biology to address global problems: synthetic biology markets, needs, and applications. *Industrial Biotechnology* 10(3):140–149.
23. Lee SY, et al. (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis* 2(1):18–33.
24. Pugh EL, Kates M (1994) Acylation of proteins of the archaeobacteria *Halobacterium cutirubrum* and *Methanobacterium thermoautotrophicum*. *Biochim Biophys Acta* 1196(1):38–44.
25. Maier T, Jenni S, Ban N (2006) Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* 311(5765):1258–1262.
26. Jenni S, Leibundgut M, Maier T, Ban N (2006) Architecture of a fungal fatty acid synthase at 5 Å resolution. *Science* 311(5765):1263–1267.
27. Janßen HJ, Steinbüchel A (2014) Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnol Biofuels* 7(1):7.
28. Cho H, Cronan JE (1995) Defective export of a periplasmic enzyme disrupts regulation of fatty acid synthesis. *J Biol Chem* 270(9):4216–4219.
29. Kaneda T (1991) Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev* 55(2):288–302.
30. Choi KH, Heath RJ, Rock CO (2000) beta-ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J Bacteriol* 182(2):365–370.
31. Howard TP, et al. (2013) Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proc Natl Acad Sci U S A* 110(19):7636–7641.
32. Haushalter RW, et al. (2014) Production of anteiso-branched fatty acids in *Escherichia coli*; next generation biofuels with improved cold-flow properties. *Metab Eng* 26:111–118.

33. Bentley GJ, Jiang W, Guamán LP, Xiao Y, Zhang F (2016) Engineering *Escherichia coli* to produce branched-chain fatty acids in high percentages. *Metab Eng* 38:148–158.
34. Steen EJ, et al. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463(7280):559–562.
35. Choi YJ, Lee SY (2013) Microbial production of short-chain alkanes. *Nature* 502(7472):571–574.
36. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* 329(5991):559–562.
37. Sorigué D, et al. (2017) An algal photoenzyme converts fatty acids to hydrocarbons. *Science* 357(6354):903–907.
38. Goh E-B, Baidoo EEK, Keasling JD, Beller HR (2012) Engineering of bacterial methyl ketone synthesis for biofuels. *Appl Environ Microbiol* 78(1):70–80.
39. Goh E-B, Chen Y, Petzold CJ, Keasling JD, Beller HR (2018) Improving methyl ketone production in *Escherichia coli* by heterologous expression of NADH-dependent FabG. *Biotechnol Bioeng* 115(5):1161–1172.
40. Goh E-B, et al. (2014) Substantial improvements in methyl ketone production in *E. coli* and insights on the pathway from in vitro studies. *Metab Eng* 26:67–76.
41. Dong J, et al. (2019) Methyl ketone production by *Pseudomonas putida* is enhanced by plant-derived amino acids. *Biotechnol Bioeng* 116(8):1909–1922.
42. Yuzawa S, et al. (2017) Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth Biol* 6(1):139–147.
43. Yuzawa S, et al. (2018) Short-chain ketone production by engineered polyketide synthases in *Streptomyces albus*. *Nat Commun* 9(1):4569.
44. Withers ST, Gottlieb SS, Lieu B, Newman JD, Keasling JD (2007) Identification of isopentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. *Appl Environ Microbiol* 73(19):6277–6283.
45. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD (2012) Microbial engineering for the production of advanced biofuels. *Nature* 488(7411):320–328.
46. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* 21(7):796–802.
47. Westfall PJ, et al. (2012) Production of amorphaadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci U S A* 109(3):E111–8.
48. Zargar A, et al. (2017) Leveraging microbial biosynthetic pathways for the generation of “drop-in” biofuels. *Curr Opin Biotechnol* 45:156–163.

49. Jenke-Kodama H, Sandmann A, Müller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* 22(10):2027–2039.
50. Weber T, et al. (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43(W1):W237–W243.
51. Menendez-Bravo S, et al. (2016) High cell density production of multimethyl-branched long-chain esters in *Escherichia coli* and determination of their physicochemical properties. *Biotechnol Biofuels* 9:215.
52. Olano C, et al. (2004) Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: insights into nitrile formation. *Mol Microbiol* 52(6):1745–1756.
53. Yuzawa S, Eng CH, Katz L, Keasling JD (2013) Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase. *Biochemistry* 52(22):3791–3793.
54. Süßmuth RD, Mainz A (2017) Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew Chem Int Ed Engl* 56(14):3770–3821.
55. Ortega MA, van der Donk WA (2016) New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chemical Biology* 23(1):31–44.
56. Cooke TF, et al. (2017) Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell* 171(2):427–439.e21.
57. Davies J (2006) Where have All the Antibiotics Gone? *Can J Infect Dis Med Microbiol* 17(5):287–290.
58. Bentley SD, et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417(6885):141–147.
59. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF (2018) Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558(7710):440–444.
60. Demain AL, Fang A (2000) The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* 69:1–39.
61. Newman DJ, Cragg GM (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* 79(3):629–661.
62. Jenke-Kodama H, Dittmann E (2009) Evolution of metabolic diversity: insights from microbial polyketide synthases. *Phytochemistry* 70(15-16):1858–1866.
63. Nicholson TP, et al. (2003) First in vitro directed biosynthesis of new compounds by a minimal type II polyketide synthase: evidence for the mechanism of chain length determination. *Chem Commun (Camb)* (6):686–687.

64. Tang Y, Lee TS, Khosla C (2004) Engineered biosynthesis of regioselectively modified aromatic polyketides using bimodular polyketide synthases. *PLoS Biol* 2(2):E31.
65. Xu Z, Schenk A, Hertweck C (2007) Molecular analysis of the benastatin biosynthetic pathway and genetic engineering of altered fatty acid-polyketide hybrids. *J Am Chem Soc* 129(18):6022–6030.
66. Lee TS, Khosla C, Tang Y (2005) Engineered biosynthesis of aklanonic acid analogues. *J Am Chem Soc* 127(35):12254–12262.
67. Bisang C, et al. (1999) A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* 401(6752):502–505.
68. Tang Y, Tsai S-C, Khosla C (2003) Polyketide chain length control by chain length factor. *J Am Chem Soc* 125(42):12708–12709.
69. Shen Y, et al. (1999) Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes. *Proc Natl Acad Sci U S A* 96(7):3622–3627.
70. Carreras CW, Khosla C (1998) Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. *Biochemistry* 37(8):2084–2088.
71. Pel HJ, et al. (2007) Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol* 25(2):221–231.
72. Cox RJ, Simpson TJ (2009) Fungal type I polyketide synthases. *Meth Enzymol* 459:49–78.
73. Chooi Y-H, Tang Y (2012) Navigating the fungal polyketide chemical space: from genes to molecules. *J Org Chem* 77(22):9933–9953.
74. Cox RJ (2007) Polyketides, proteins and genes in fungi: programmed nanomachines begin to reveal their secrets. *Org Biomol Chem* 5(13):2010–2026.
75. Crawford JM, Townsend CA (2010) New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* 8(12):879–889.
76. Birch AJ, Donovan FW (1953) Studies in relation to biosynthesis. *Australian Journal of Chemistry*.
77. Bingle LE, Simpson TJ, Lazarus CM (1999) Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. *Fungal Genet Biol* 26(3):209–223.
78. Nicholson TP, et al. (2001) Design and utility of oligonucleotide gene probes for fungal polyketide synthases. *Chem Biol* 8(2):157–178.
79. Crawford JM, et al. (2008) Deconstruction of iterative multidomain polyketide synthase function. *Science* 320(5873):243–246.

80. Crawford JM, Dancy BCR, Hill EA, Udvary DW, Townsend CA (2006) Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc Natl Acad Sci U S A* 103(45):16728–16733.
81. Austin MB, et al. (2004) Crystal structure of a bacterial type III polyketide synthase and enzymatic control of reactive polyketide intermediates. *J Biol Chem* 279(43):45162–45174.
82. Herbst DA, et al. (2018) The structural organization of substrate loading in iterative polyketide synthases. *Nat Chem Biol* 14(5):474–479.
83. Crawford JM, et al. (2009) Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* 461(7267):1139–1143.
84. Barajas JF, et al. (2017) Polyketide mimetics yield structural and mechanistic insights into product template domain function in nonreducing polyketide synthases. *Proc Natl Acad Sci U S A* 114(21):E4142–E4148.
85. Li Y, et al. (2010) Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J Biol Chem* 285(30):22764–22773.
86. Korman TP, et al. (2010) Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci U S A* 107(14):6246–6251.
87. Katz L (2009) The DEBS paradigm for type I modular polyketide synthases and beyond. *Meth Enzymol* 459:113–142.
88. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* 252(5006):675–679.
89. Cane DE (2010) Programming of erythromycin biosynthesis by a modular polyketide synthase. *J Biol Chem* 285(36):27517–27523.
90. Helfrich EJM, Piel J (2016) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* 33(2):231–316.
91. Piel J (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc Natl Acad Sci U S A* 99(22):14002–14007.
92. Bretschneider T, et al. (2013) Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* 502(7469):124–128.
93. Pöplau P, Frank S, Morinaka BI, Piel J (2013) An enzymatic domain for the formation of cyclic ethers in complex polyketides. *Angew Chem Int Ed Engl* 52(50):13215–13218.

94. Poust S, Hagen A, Katz L, Keasling JD (2014) Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Curr Opin Biotechnol* 30:32–39.
95. Liu Q, et al. (2015) Engineering an iterative polyketide pathway in *Escherichia coli* results in single-form alkene and alkane overproduction. *Metab Eng* 28:82–90.
96. Hagen A, et al. (2016) Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth Biol* 5(1):21–27.
97. Menendez-Bravo S, Comba S, Sabatini M, Arabolaza A, Gramajo H (2014) Expanding the chemical diversity of natural esters by engineering a polyketide-derived pathway into *Escherichia coli*. *Metab Eng* 24:97–106.
98. Guo X, Liu T, Valenzano CR, Deng Z, Cane DE (2010) Mechanism and stereospecificity of a fully saturating polyketide synthase module: nanchangmycin synthase module 2 and its dehydratase domain. *J Am Chem Soc* 132(42):14694–14696.
99. Barajas JF, Blake-Hedges JM, Bailey CB, Curran S, Keasling JD (2017) Engineered polyketides: Synergy between protein and host level engineering. *Synthetic and Systems Biotechnology* 2(3):147–166.
100. Ray L, Moore BS (2016) Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Nat Prod Rep* 33(2):150–161.
101. Long PF, et al. (2002) Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol Microbiol* 43(5):1215–1225.
102. Oefner C, Schulz H, D'Arcy A, Dale GE (2006) Mapping the active site of *Escherichia coli* malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallogr D Biol Crystallogr* 62(Pt 6):613–618.
103. Dutton CJ, et al. (1991) Novel avermectins produced by mutational biosynthesis. *J Antibiot* 44(3):357–365.
104. Moss SJ, et al. (2006) Biosynthesis of the angiogenesis inhibitor borrelidin: directed biosynthesis of novel analogues. *Chem Commun (Camb)* (22):2341–2343.
105. Hagen A, et al. (2014) In vitro analysis of carboxyacyl substrate tolerance in the loading and first extension modules of borrelidin polyketide synthase. *Biochemistry* 53(38):5975–5977.
106. Wang F, et al. (2015) Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase. *ACS Chem Biol* 10(4):1017–1025.
107. Nguyen T, et al. (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* 26(2):225–233.

108. Gay DC, et al. (2014) A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure* 22(3):444–451.
109. Lohman JR, et al. (2015) Structural and evolutionary relationships of “AT-less” type I polyketide synthase ketosynthases. *Proc Natl Acad Sci U S A* 112(41):12693–12698.
110. Jenner M, et al. (2013) Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angew Chem Int Ed Engl* 52(4):1143–1147.
111. Jenner M, et al. (2015) Acyl-chain elongation drives ketosynthase substrate selectivity in trans-acyltransferase polyketide synthases. *Angew Chem Int Ed Engl* 54(6):1817–1821.
112. Zhang L, et al. (2017) Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew Chem Int Ed Engl* 56(7):1740–1745.
113. Busch B, et al. (2013) Multifactorial control of iteration events in a modular polyketide assembly line. *Angew Chem Int Ed Engl* 52(20):5285–5289.
114. Sugimoto Y, et al. (2015) Freedom and constraint in engineered noncolinear polyketide assembly lines. *Chem Biol* 22(2):229–240.
115. Wu J, Kinoshita K, Khosla C, Cane DE (2004) Biochemical analysis of the substrate specificity of the beta-ketoacyl-acyl carrier protein synthase domain of module 2 of the erythromycin polyketide synthase. *Biochemistry* 43(51):16301–16310.
116. Murphy AC, Hong H, Vance S, Broadhurst RW, Leadlay PF (2016) Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem Commun (Camb)* 52(54):8373–8376.
117. Robbins T, Kapilivsky J, Cane DE, Khosla C (2016) Roles of conserved active site residues in the ketosynthase domain of an assembly line polyketide synthase. *Biochemistry* 55(32):4476–4484.
118. Eng CH, et al. (2018) ClusterCAD: a computational platform for type I modular polyketide synthase design. *Nucleic Acids Res* 46(D1):D509–D515.
119. Weissman KJ (2006) The structural basis for docking in modular polyketide biosynthesis. *Chembiochem* 7(3):485–494.
120. Hughes AJ, Detelich JF, Keatinge-Clay AT (2012) Employing a polyketide synthase module and thioesterase in the semipreparative biocatalysis of diverse triketide pyrones. *Medchemcomm* 3(8):956.
121. Menzella HG, et al. (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol* 23(9):1171–1176.
122. Menzella HG, Carney JR, Santi DV (2007) Rational design and assembly of synthetic trimodular polyketide synthases. *Chem Biol* 14(2):143–151.

123. Keatinge-Clay AT (2017) The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chem Rev* 117(8):5334–5366.
124. Wilson MC, Moore BS (2012) Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat Prod Rep* 29(1):72–86.
125. Tang Y, Chen AY, Kim C-Y, Cane DE, Khosla C (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem Biol* 14(8):931–943.
126. Tang Y, Kim C-Y, Mathews II, Cane DE, Khosla C (2006) The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* 103(30):11124–11129.
127. Haydock SF, et al. (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett* 374(2):246–248.
128. Reeves CD, et al. (2001) Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* 40(51):15464–15470.
129. Sundermann U, et al. (2013) Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem Biol* 8(2):443–450.
130. Dunn BJ, Cane DE, Khosla C (2013) Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* 52(11):1839–1841.
131. Oliynyk M, Brown MJ, Cortés J, Staunton J, Leadlay PF (1996) A hybrid modular polyketide synthase obtained by domain swapping. *Chem Biol* 3(10):833–839.
132. Keatinge-Clay AT (2016) Stereocontrol within polyketide assembly lines. *Nat Prod Rep* 33(2):141–149.
133. Zheng J, Piasecki SK, Keatinge-Clay AT (2013) Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling α -substituent stereochemistry. *ACS Chem Biol* 8(9):1964–1971.
134. Valenzano CR, Lawson RJ, Chen AY, Khosla C, Cane DE (2009) The biochemical basis for stereochemical control in polyketide biosynthesis. *J Am Chem Soc* 131(51):18501–18511.
135. Keatinge-Clay AT (2012) The structures of type I polyketide synthases. *Nat Prod Rep* 29(10):1050–1073.
136. Zheng J, Taylor CA, Piasecki SK, Keatinge-Clay AT (2010) Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* 18(8):913–922.

137. Keatinge-Clay AT (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem Biol* 14(8):898–908.
138. Keatinge-Clay AT, Stroud RM (2006) The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* 14(4):737–748.
139. Bailey CB, Pasman ME, Keatinge-Clay AT (2016) Substrate structure-activity relationships guide rational engineering of modular polyketide synthase ketoreductases. *Chem Commun (Camb)* 52(4):792–795.
140. Reid R, et al. (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* 42(1):72–79.
141. Valenzano CR, et al. (2010) Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase. *J Am Chem Soc* 132(42):14697–14699.
142. Kandziora N, et al. (2014) Uncovering the origin of Z-configured double bonds in polyketides: intermediate E-double bond formation during borrelidin biosynthesis. *Chem Sci* 5(9):3563.
143. Gay D, You Y-O, Keatinge-Clay A, Cane DE (2013) Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry* 52(49):8916–8928.
144. Alhamadsheh MM, Palaniappan N, Daschouduri S, Reynolds KA (2007) Modular polyketide synthases and cis double bond formation: establishment of activated cis-3-cyclohexylpropenoic acid as the diketide intermediate in phoslactomycin biosynthesis. *J Am Chem Soc* 129(7):1910–1911.
145. Keatinge-Clay A (2008) Crystal structure of the erythromycin polyketide synthase dehydratase. *J Mol Biol* 384(4):941–953.
146. Akey DL, et al. (2010) Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* 18(1):94–105.
147. Kwan DH, et al. (2008) Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. *Chem Biol* 15(11):1231–1240.
148. Zheng J, Gay DC, Demeler B, White MA, Keatinge-Clay AT (2012) Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat Chem Biol* 8(7):615–621.
149. Donadio S, McAlpine JB, Sheldon PJ, Jackson M, Katz L (1993) An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc Natl Acad Sci U S A* 90(15):7119–7123.
150. McDaniel R, et al. (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc Natl Acad Sci U S A* 96(5):1846–1851.

151. Gaisser S, et al. (2003) Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of *Streptomyces avermitilis* ATCC31272. *Org Biomol Chem* 1(16):2840–2847.
152. Du L, Lou L (2010) PKS and NRPS release mechanisms. *Nat Prod Rep* 27(2):255–278.
153. Horsman ME, Hari TPA, Boddy CN (2016) Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat Prod Rep* 33(2):183–202.
154. Pinto A, Wang M, Horsman M, Boddy CN (2012) 6-Deoxyerythronolide B synthase thioesterase-catalyzed macrocyclization is highly stereoselective. *Org Lett* 14(9):2278–2281.
155. Kao CM, Luo G, Katz L, Cane DE, Khosla C (1996) Engineered biosynthesis of structurally diverse tetraketides by a trimodular polyketide synthase. *J Am Chem Soc* 118(38):9184–9185.
156. Kao CM, et al. (1997) Gain of Function Mutagenesis of the Erythromycin Polyketide Synthase. 2. Engineered Biosynthesis of an Eight-Membered Ring Tetraketide Lactone. *J Am Chem Soc* 119(46):11339–11340.
157. Kao CM, Luo G, Katz L, Cane DE, St CK (1995) Manipulation of Macrolide Ring Size by Directed Mutagenesis of a Modular Polyketide Synthase. *J Am Chem Soc* (16):9105–9106.
158. Jacobsen JR, Hutchinson CR, Cane DE, Khosla C (1997) Precursor-directed biosynthesis of erythromycin analogs by an engineered polyketide synthase. *Science* 277(5324):367–369.
159. Wang M, Boddy CN (2008) Examining the role of hydrogen bonding interactions in the substrate specificity for the loading step of polyketide synthase thioesterase domains. *Biochemistry* 47(45):11793–11803.
160. Wang M, Opare P, Boddy CN (2009) Polyketide synthase thioesterases catalyze rapid hydrolysis of peptidyl thioesters. *Bioorg Med Chem Lett* 19(5):1413–1415.
161. Lu H, Tsai S-C, Khosla C, Cane DE (2002) Expression, Site-Directed Mutagenesis, and Steady State Kinetic Analysis of the Terminal Thioesterase Domain of the Methymycin/Picromycin Polyketide Synthase[†]. *Biochemistry* 41(42):12590–12597.
162. Sharma KK, Boddy CN (2007) The thioesterase domain from the pimaricin and erythromycin biosynthetic pathways can catalyze hydrolysis of simple thioester substrates. *Bioorg Med Chem Lett* 17(11):3034–3037.
163. Barajas JF, et al. (2015) Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols. *Chem Biol* 22(8):1018–1029.

164. Gehret JJ, et al. (2011) Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J Biol Chem* 286(16):14445–14454.
165. Khosla C, Gokhale RS, Jacobsen JR, Cane DE (1999) Tolerance and specificity of polyketide synthases. *Annu Rev Biochem* 68:219–253.
166. Wu N, Tsuji SY, Cane DE, Khosla C (2001) Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J Am Chem Soc* 123(27):6465–6474.
167. Klaus M, et al. (2016) Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *J Biol Chem* 291(31):16404–16415.
168. Tsuji SY, Cane DE, Khosla C (2001) Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* 40(8):2326–2331.
169. Chandran SS, Menzella HG, Carney JR, Santi DV (2006) Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of ketosynthase domains. *Chem Biol* 13(5):469–474.
170. Alekseyev VY, Liu CW, Cane DE, Puglisi JD, Khosla C (2007) Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci* 16(10):2093–2107.
171. Miyanaga A, Iwasawa S, Shinohara Y, Kudo F, Eguchi T (2016) Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc Natl Acad Sci U S A* 113(7):1802–1807.
172. Whicher JR, et al. (2014) Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* 510(7506):560–564.
173. Dutta S, et al. (2014) Structure of a modular polyketide synthase. *Nature* 510(7506):512–517.
174. Weissman KJ (2015) The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* 11(9):660–670.
175. Olano C, et al. (2004) Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: cluster analysis and assignment of functions. *Chem Biol* 11(1):87–97.
176. Olano C, et al. (2003) Evidence from engineered gene fusions for the repeated use of a module in a modular polyketide synthase. *Chem Commun (Camb)* (22):2780–2782.
177. Chen H, Du L (2016) Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl Microbiol Biotechnol* 100(2):541–557.

178. Wilkinson B, et al. (2000) Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chem Biol* 7(2):111–117.
179. Hardt IH, et al. (2001) New Natural Epothilones from *Sorangium cellulosum*, Strains So ce90/B2 and So ce90/D13: Isolation, Structure Elucidation, and SAR Studies. *J Nat Prod* 64(7):847–856.
180. Beck BJ, Aldrich CC, Fecik RA, Reynolds KA, Sherman DH (2003) Iterative chain elongation by a pikromycin monomodular polyketide synthase. *J Am Chem Soc* 125(16):4682–4683.
181. Busch B, Ueberschaar N, Sugimoto Y, Hertweck C (2012) Interchain retrotransfer of aureothin intermediates in an iterative polyketide synthase module. *J Am Chem Soc* 134(30):12382–12385.
182. Gaitatzis N, et al. (2002) The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J Biol Chem* 277(15):13082–13090.
183. He J, Hertweck C (2003) Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem Biol* 10(12):1225–1232.
184. Traitcheva N, Jenke-Kodama H, He J, Dittmann E, Hertweck C (2007) Non-colinear polyketide biosynthesis in the aureothin and neo-aureothin pathways: an evolutionary perspective. *Chembiochem* 8(15):1841–1849.
185. Mochizuki S, et al. (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol* 48(6):1501–1510.
186. Menche D, et al. (2008) Stereochemical determination and complex biosynthetic assembly of etnangien, a highly potent RNA polymerase inhibitor from the myxobacterium *Sorangium cellulosum*. *J Am Chem Soc* 130(43):14234–14243.
187. Meiser P, et al. (2008) DKxanthene biosynthesis--understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism. *Chem Biol* 15(8):771–781.
188. Müller S, et al. (2014) Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chem Biol* 21(7):855–865.
189. Hong H, et al. (2016) Evidence for an iterative module in chain elongation on the azalomycin polyketide synthase. *Beilstein J Org Chem* 12(1):2164–2172.
190. Kuo J, Lynch SR, Liu CW, Xiao X, Khosla C (2016) Partial In Vitro Reconstitution of an Orphan Polyketide Synthase Associated with Clinical Cases of Nocardiosis. *ACS Chem Biol* 11(9):2636–2641.

191. Low ZJ, et al. (2018) Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment. *Sci Rep* 8(1):1594.
192. Val D, Banu G, Seshadri K, Lindqvist Y, Dehesh K (2000) Re-engineering ketoacyl synthase specificity. *Structure* 8(6):565–566.
193. Keatinge-Clay AT, Maltby DA, Medzihradzsky KF, Khosla C, Stroud RM (2004) An antibiotic factory caught in action. *Nat Struct Mol Biol* 11(9):888–893.
194. McDaniel R, Ebert-Khosla S, Hopwood DA, Khosla C (1993) Engineered biosynthesis of novel polyketides. *Science* 262(5139):1546–1550.
195. Ma SM, et al. (2007) Enzymatic synthesis of aromatic polyketides using PKS4 from *Gibberella fujikuroi*. *J Am Chem Soc* 129(35):10642–10643.
196. Zhang W, Li Y, Tang Y (2008) Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*. *Proc Natl Acad Sci U S A* 105(52):20683–20688.
197. Wattana-amorn P, et al. (2010) Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry* 49(10):2186–2193.
198. Campuzano ID, Shoolingin-Jordan PM (1998) Incubation of 6-methylsalicylic acid synthase with alternative starter units in the absence of NADPH and the identification of the resulting triaceticacid lactones. *Biochem Soc Trans* 26(3):S284.
199. Ma SM, et al. (2009) Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* 326(5952):589–592.
200. Kapust RB, Waugh DS (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8(8):1668–1674.
201. Quadri LE, et al. (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 37(6):1585–1595.
202. Dorrestein PC, et al. (2006) Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* 45(42):12756–12766.
203. Moss SJ, Martin CJ, Wilkinson B (2004) Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat Prod Rep* 21(5):575–593.
204. Tang L, et al. (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science* 287(5453):640–642.
205. Lowry B, et al. (2013) In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* 135(45):16809–16812.

206. Katiyar SS, Cleland WW, Porter JW (1975) Fatty acid synthetase. A steady state kinetic analysis of the reaction catalyzed by the enzyme from pigeon liver. *J Biol Chem* 250(7):2709–2717.
207. Kotowska M, Pawlik K (2014) Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl Microbiol Biotechnol* 98(18):7735–7746.
208. Cacho RA, et al. (2015) Understanding programming of fungal iterative polyketide synthases: the biochemical basis for regioselectivity by the methyltransferase domain in the lovastatin megasynthase. *J Am Chem Soc* 137(50):15688–15691.
209. Huitt-Roehl CR, et al. (2015) Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA. *ACS Chem Biol* 10(6):1443–1449.
210. Hillson NJ, Rosengarten RD, Keasling JD (2012) j5 DNA assembly design automation software. *ACS Synth Biol* 1(1):14–21.
211. Tropea JE, Cherry S, Waugh DS (2009) Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol* 498:297–307.
212. Lee TS, et al. (2011) BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* 5:12.
213. Chen J, Densmore D, Ham TS, Keasling JD, Hillson NJ (2012) DeviceEditor visual biological CAD canvas. *J Biol Eng* 6(1):1.
214. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.
215. Tae H, Sohng JK, Park K (2009) Development of an analysis program of type I polyketide synthase gene clusters using homology search and profile hidden Markov model. *J Microbiol Biotechnol* 19(2):140–146.
216. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
217. Wessel D, Flügge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 138(1):141–143.
218. MacLean B, et al. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7):966–968.
219. Sharma V, et al. (2018) Panorama public: A public repository for quantitative data sets processed in skyline. *Mol Cell Proteomics* 17(6):1239–1244.
220. Walsh CT, Gehring AM, Weinreb PH, Quadri LE, Flugel RS (1997) Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr Opin Chem Biol* 1(3):309–315.

221. Heathcote ML, Staunton J, Leadlay PF (2001) Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem Biol* 8(2):207–220.
222. Zhou Y, et al. (2008) Selective removal of aberrant extender units by a type II thioesterase for efficient FR-008/candicidin biosynthesis in *Streptomyces* sp. strain FR-008. *Appl Environ Microbiol* 74(23):7235–7242.
223. Kotowska M, Pawlik K, Smulczyk-Krawczynszyn A, Bartosz-Bechowski H, Kuczek K (2009) Type II thioesterase ScoT, associated with *Streptomyces coelicolor* A3(2) modular polyketide synthase Cpk, hydrolyzes acyl residues and has a preference for propionate. *Appl Environ Microbiol* 75(4):887–896.
224. Kim BS, Cropp TA, Beck BJ, Sherman DH, Reynolds KA (2002) Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. *J Biol Chem* 277(50):48028–48034.
225. Liu T, Lin X, Zhou X, Deng Z, Cane DE (2008) Mechanism of thioesterase-catalyzed chain release in the biosynthesis of the polyether antibiotic nanchangmycin. *Chem Biol* 15(5):449–458.
226. Yeh E, Kohli RM, Bruner SD, Walsh CT (2004) Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem* 5(9):1290–1293.
227. Koglin A, et al. (2008) Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* 454(7206):907–911.
228. Schwarzer D, Mootz HD, Linne U, Marahiel MA (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* 99(22):14083–14088.
229. Claxton HB, Akey DL, Silver MK, Admiraal SJ, Smith JL (2009) Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *J Biol Chem* 284(8):5021–5029.
230. Whicher JR, et al. (2011) Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in *Streptomyces coelicolor*. *J Biol Chem* 286(25):22558–22569.
231. Guntaka NS, Healy AR, Crawford JM, Herzon SB, Bruner SD (2017) Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway. *ACS Chem Biol* 12(10):2598–2608.
232. Cantu DC, Chen Y, Reilly PJ (2010) Thioesterases: a new perspective based on their primary and tertiary structures. *Protein Sci* 19(7):1281–1295.
233. Franke J, Hertweck C (2016) Biomimetic thioesters as probes for enzymatic assembly lines: synthesis, applications, and challenges. *Cell Chemical Biology* 23(10):1179–1192.

234. Ellman GL (1958) A colorimetric method for determining low concentrations of mercaptans. *Arch Biochem Biophys* 74(2):443–450.
235. Ollis DL, et al. (1992) The α / β hydrolase fold. *Protein Eng Des Sel* 5(3):197–211.
236. Tsai SC, et al. (2001) Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc Natl Acad Sci U S A* 98(26):14808–14813.
237. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* 105(4):962–974.
238. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* 44(W1):W424–9.
239. Bruner SD, et al. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* 10(3):301–310.
240. Tsai S-C, Lu H, Cane DE, Khosla C, Stroud RM (2002) Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* 41(42):12598–12606.
241. Nardini M, Dijkstra BW (1999) α/β Hydrolase fold enzymes: the family keeps growing. *Current opinion in structural biology* 9(6):732–737.
242. Zhang W, et al. (2011) Crystal structure of FAS thioesterase domain with polyunsaturated fatty acyl adduct and inhibition by dihomo-gamma-linolenic acid. *Proc Natl Acad Sci U S A* 108(38):15757–15762.
243. Giraldes JW, et al. (2006) Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat Chem Biol* 2(10):531–536.
244. Frueh DP, et al. (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* 454(7206):903–906.
245. Liu Y, Zheng T, Bruner SD (2011) Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem Biol* 18(11):1482–1488.
246. Huguenin-Dezot N, et al. (2019) Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* 565(7737):112–117.
247. Jochens H, et al. (2011) Protein engineering of α/β -hydrolase fold enzymes. *Chembiochem* 12(10):1508–1517.
248. Chakravarty B, Gu Z, Chirala SS, Wakil SJ, Quioco FA (2004) Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain. *Proc Natl Acad Sci U S A* 101(44):15567–15572.

249. Ziesack M, et al. (2018) Chimeric Fatty Acyl-Acyl Carrier Protein Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression. *Appl Environ Microbiol* 84(10).
250. Hari TPA, Labana P, Boileau M, Boddy CN (2014) An evolutionary model encompassing substrate specificity and reactivity of type I polyketide synthase thioesterases. *Chembiochem* 15(18):2656–2661.
251. De los Santos EL, Challis GL (2017) clusterTools: Proximity Searches For Functional Elements To Identify Putative Biosynthetic Gene Clusters. *BioRxiv*.
252. Medema MH, et al. (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* 11(9):625–631.
253. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
254. Curran SC, et al. (2018) Probing the Flexibility of an Iterative Modular Polyketide Synthase with Non-Native Substrates *in Vitro*. *ACS Chem Biol* 13(8):2261–2268.
255. Classen S, et al. (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* 46(Pt 1):1–13.
256. Dyer KN, et al. (2014) High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol Biol* 1091:245–258.
257. Hura GL, et al. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* 6(8):606–612.
258. Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496(7446):477–481.
259. Jancarik J, Kim SH (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Crystallogr* 24(4):409–411.
260. Pereira JH, McAndrew RP, Tomaleri GP, Adams PD (2017) Berkeley Screen: a set of 96 solutions for general macromolecular crystallization. *J Appl Crystallogr* 50(Pt 5):1352–1358.
261. Winter G, Lobley CMC, Prince SM (2013) Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1260–1273.
262. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658–674.
263. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
264. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352–367.

265. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
266. Davis IW, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375–83.
267. Vandamme EJ, Soetaert W (2002) Bioflavours and fragrances via fermentation and biocatalysis. *Journal of Chemical Technology & Biotechnology* 77(12):1323–1332.
268. Müller J, et al. (2013) Engineering of *Ralstonia eutropha* H16 for autotrophic and heterotrophic production of methyl ketones. *Appl Environ Microbiol* 79(14):4433–4439.
269. Hanko EKR, et al. (2018) Engineering β -oxidation in *Yarrowia lipolytica* for methyl ketone production. *Metab Eng* 48:52–62.
270. Kornberg A, Ochoa S, Mehler AH (1948) Spectrophotometric studies on the decarboxylation of beta-keto acids. *J Biol Chem* 174(1):159–172.
271. Liu H, Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* 8:91.
272. Yuzawa S, et al. (2017) Heterologous Gene Expression of N-Terminally Truncated Variants of LipPks1 Suggests a Functionally Critical Structural Motif in the N-terminus of Modular Polyketide Synthase. *ACS Chem Biol* 12(11):2725–2729.
273. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* 291(5509):1790–1792.
274. Mutka SC, Bondi SM, Carney JR, Da Silva NA, Kealey JT (2006) Metabolic pathway engineering for complex polyketide biosynthesis in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 6(1):40–47.
275. Liu R, Deng Z, Liu T (2018) *Streptomyces* species: Ideal chassis for natural product discovery and overproduction. *Metab Eng* 50:74–84.
276. Pfeifer BA, Khosla C (2001) Biosynthesis of polyketides in heterologous hosts. *Microbiol Mol Biol Rev* 65(1):106–118.
277. Bihlmaier C, et al. (2006) Biosynthetic gene cluster for the polyenoyltetramic acid alpha-lipomycin. *Antimicrob Agents Chemother* 50(6):2113–2121.
278. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190.