UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Understanding and Engineering of the Biosynthesis of Terminal Alkene and Alkyne Containing Natural Products

Permalink

https://escholarship.org/uc/item/2j69s413

Author

Gu, Di

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Understanding and Engineering of the Biosynthesis of Terminal Alkene and Alkyne Containing Natural Products

By

Di Gu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Wenjun Zhang, Co-Chair

Professor Michelle C. Chang, Co-Chair

Professor Michael A. Marletta

Professor John E. Dueber

Spring 2024

Understanding and Engineering of the Biosynthesis of Terminal Alkene and Alkyne Containing Natural Products

Copyright © 2024

by Di Gu

Abstract

Understanding and Engineering of the Biosynthesis of Terminal Alkene and Alkyne Containing Natural Products

by

Di Gu

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Wenjun Zhang, Co-Chair
Professor Michelle C. Chang, Co-Chair

Natural products have played a critical role in various fields for decades. Studying natural product biosynthesis facilitates the discovery of novel bioactive compounds and innovates engineering strategies to create molecules with specific scaffolds or functional groups. In addition to producing tailored molecules, biosynthetic engineering can also be used to achieve mass production of useful compounds. Among all kinds of natural products, terminal alkene and alkyne containing compounds have garnered significant attention due to the reactivity of their terminal functionalities and broad applications in chemical biology research and pharmaceutical industry. Therefore, it is important to understand the biosynthesis of terminal alkene and alkyne containing natural products and engineer pathways to leverage the biosynthetic machinery.

Salivabactin is a terminal alkene containing natural product isolated from *Streptococcus salivarius* K12, which exhibits potent antimicrobial activities towards gram-positive bacteria. Considering its bioactivity and unavailability though synthesis, it is necessary to understand the biosynthesis of salivabactin. Based on enzymatic assays *in vitro*, we elucidated the biosynthesis and uncovered the minimal set of enzymes required for salivabactin production. Notably, an unprecedented thioesterase involved in the terminal alkene formation was characterized by site-directed mutagenesis and structural elucidation. With the understanding of biosynthesis, we were able to engineer the pathway to increase compound titer. Through heterologous expression and growth optimization in *E. coli*, we achieved a remarkable seven-fold enhancement in titer compared to that in the native host. This work broadens our understanding of terminal alkene biosynthesis and facilitates the mass production of new antibiotics to combat emerging multi-drug resistance.

In addition to terminal alkene containing natural products, there are also diverse natural products with the terminal alkyne moiety, formed through various biosynthetic mechanisms. Among them, we focused on the family of membrane-bound desaturases/acetylenases, represented by JamB and TtuB. These enzymes are capable of catalyzing the terminal alkyne formation on acyl carrier protein (ACP)-tethered substrates. Intriguing mechanistic questions surrounding JamB and TtuB prompted the biochemical and structural studies, necessitating solubilized and active forms of these membrane-bound enzymes. To achieve this, we employed different strategies, including residue swapping, cell-free protein synthesis and lipoprotein shielding. Our efforts

yielded promising results in terms of solubilization and enzyme activity, paving the way for further mechanistic exploration into terminal alkyne biosynthesis. Meanwhile, with the successful engineering to generate terminal alkyne tagged polyketides, we extended the applicability of terminal alkyne tagging to lipopeptides, as demonstrated by two NRPS systems *in vivo* and *in vitro*. Notably, we employed an innovative engineering strategy involving a non-elongating ketosynthase domain, which facilitated the transfer of acyl chains between ACPs. This establishes a versatile method for incorporating non-cognate ACPs into assembly lines.

Table of Contents

Table of Contents ..... **i**

List of Figures ..... **iii**

List of Tables ..... **v**

Acknowledgements ..... **vi**

Chapter 1. Introduction **1**

1.1 Natural Product Biosynthesis 1

1.2 Biosynthetic Pathway Engineering 3

1.3 Terminal Alkene Containing Natural Products and Biosynthesis 5

1.4 Terminal Alkyne Containing Natural Products and Biosynthesis 7

Chapter 2. Biosynthesis of Salivabactin and Characterization of Its Biosynthetic Enzymes **11**

2.1 Introduction 11

2.2 Results 14

2.2.1 Proposed biosynthetic pathway 14

2.2.2 Elucidation of salivabactin biosynthesis and in vitro reconstitution 15

2.2.3 Biochemical studies on TE domain and other biosynthetic enzymes 18

2.2.4 Structural biology studies on TE domain and proposed releasing mechanism 21

2.3 Discussion 24

2.4 Materials and Methods 26

Chapter 3. Biosynthetic Pathway Engineering of Salivabactin **33**

3.1 Introduction 33

3.2 Results 34

3.2.1 Heterologous expression of sar BGC in E. coli 34

3.2.2 Optimization of growth conditions 37

3.2.3 Starter unit engineering 39

3.3 Discussion 40

3.4 Materials and Methods 41

Chapter 4. Biochemical Studies of Terminal Alkyne Formation Enzymes 45

4.1 Introduction 45

4.2 Results 46

4.2.1 Solubilization by QTY code and biochemical characterization 47

4.2.2 Solubilization by CFPS with RHP and biochemical characterization 49

4.2.3 Solubilization by SIMPLEx and biochemical characterization 52

4.2.4 Structural modeling of JamB and mechanistic insights 54

4.3 Discussion 55

4.4 Materials and Methods 56

Chapter 5. Modular Assembly Line Engineering of Terminal Alkyne Tagged Natural Products 61

5.1 Introduction 61

5.2 Results 65

5.2.1 Biosynthetic engineering in vivo in EpnG system 65

5.2.2 Biosynthetic engineering in vitro in MmaA system 67

5.3 Discussion 70

5.4 Materials and Methods 71

Chapter 6. Conclusion 75

References 78

Appendices 87

List of Figures

Figure 1-1. Examples of natural products with pharmaceutical applications 1
Figure 1-2. Enzymatic machinery of PKSs and NRPSs. 2
Figure 1-3. Schematic of four biosynthetic pathway engineering strategies 3
Figure 1-4. Examples of terminal alkene containing natural products 6
Figure 1-5. Examples of biosynthetic enzymes of terminal alkene formation in natural products 7
Figure 1-6. Examples of terminal alkyne containing natural products 8
Figure 1-7. Known biosynthetic enzymes for terminal alkyne formation 9
Figure 2-1. Discovery of a new antibiotic from SAL 11
Figure 2-2. Chemical structure of the two salivabactin isomers, salivabactin A and salivabactin B 12
Figure 2-3. Antimicrobial activities of salivabactin 13
Figure 2-4. Proposed biosynthetic pathway of salivabactin 15
Figure 2-5. SDS-PAGE gel of each biosynthetic enzyme 16
Figure 2-6. Biosynthetic characterization of SarD, SarE and SarJ 16
Figure 2-7. Time course results of the production of **(A)** salivabactin and **(B)** intermediate-335 17
Figure 2-8. *In vitro* reconstitution of salivabactin biosynthesis 18
Figure 2-9. SDS-PAGE gel of different versions of TE in SarH 19
Figure 2-10. Biochemical characterization of different types of TE in SarH 20
Figure 2-11. SDS-PAGE gel of SarH without ACP and TE mutants in SarH 21
Figure 2-12. Crystal structure of TE in SarH 22
Figure 2-13. Sequence alignment of SarH_TE with structural homologues 22
Figure 2-14. Enzymatic assays of TE mutants, measured by LC/HRMS-EIC 23
Figure 2-15. Docking model of intermediate-335 covalently linked with C137 in SarH_TE 24
Figure 2-16. Proposed mechanism for the formation of terminal alkene by TE domain in SarH 24
Figure 2-17. Alignment of SarH_TE and YbfF 25
Figure 3-1. Time course for salivabactin production 33
Figure 3-2. Schematic of four strains for heterologous expression of *sar* BGC 35
Figure 3-3. Comparison of salivabactin production from four strains 36
Figure 3-4. Time course for salivabactin production from strain **1** in TB medium 37
Figure 3-5. Impact of substrate feeding on salivabactin production 38
Figure 3-6. Titer comparison of salivabactin from different hosts and different growth conditions 39
Figure 3-7. Results of *in vivo* and *in vitro* starter unit engineering 40
Figure 4-1. Solubilization of JamB by QTY code engineering 47
Figure 4-2. Activity tests of JamB(QTY) *in vivo* and *in vitro* 48
Figure 4-3. First round of CFPS optimization 50
Figure 4-4. Second round of CFPS optimization 50

Figure 4-5. Third round of CFPS optimization 51
Figure 4-6. Purified JamB-sfGFP and TtuB from CFPS 52
Figure 4-7. SDS-PAGE gel of solubilized JamB/TtuB with SIMPLEx method..... 53
Figure 4-8. Activity tests of MBP-TtuB-ApoAI* *in vivo* 54
Figure 4-9. Proposed JamB structural model 55
Figure 5-1. Type I PKS starter unit engineering for terminal alkyne tagged polyketide biosynthesis 62
Figure 5-2. Proposed biosynthesis of a peptidyl epoxyketone produced in *E. coli* 64
Figure 5-3. Proposed biosynthesis of an isonitrile lipopeptide produced in *E. coli* 65
Figure 5-4. Proposed biosynthesis of terminal alkyne tagged epoxyketone metabolites in *E. coli* 66
Figure 5-5. Titer comparison of compounds in extracts from supernatant of *E. coli* cultures with the co-expression of *ttuABC*, *epnG*, *epxE* and *epnF* 67
Figure 5-6. Biosynthesis of lipopeptides with MmaA *in vitro* 68
Figure 5-7. Enzymatic assay results of compound **15** production 69
Figure 5-8. Enzymatic assay results of compound **16** production 70

List of Tables

Table 2-1. MIC of salivabactin against different bacterial pathogens 13
Table 2-2. Deduced roles of open reading frames in the *sar* BGC 14
Table 2-3. Oligonucleotides used in this study 26
Table 2-4. Plasmids used in this study 28
Table 2-5. Bacterial strains used in this study 29
Table 3-1. Oligonucleotides used in this study 42
Table 3-2. Plasmids used in this study 42
Table 3-3. Bacterial strains used in this study 42
Table 4-1. Oligonucleotides used in this study 56
Table 4-2. Plasmids used in this study 58
Table 5-1. Oligonucleotides used in this study 71
Table 5-2. Plasmids used in this study 72

Acknowledgements

I would like to express my sincere gratitude to my advisor, Prof. Wenjun Zhang, for supporting and mentoring me for the past five years. Without her guidance, all the work presented in this dissertation would not have been achieved. Wenjun initiated my journey into the field of biosynthesis, igniting my passion for enzymology and biosynthetic engineering, and imparted to me the methodology of scientifically investigating problems. Not only did she offer guidance on overall scientific perspective and project management, but she also personally mentored me in project details such as experimental design and skills. She provided academic resources, respected my thoughts, and supported my professional career development. Under her guidance, I had opportunities to develop leadership and presentation skills. I am truly grateful to be one of her students, and if given the chance to make the choice again, I would unquestionably choose this lab once more.

I am deeply grateful to all the past and current members of the Zhang lab for fostering such a warm and supportive familial environment. In particular, I would like to extend my gratitude to the following members. Thank Jeffrey Li for his attentive mentorship during my summer research as an undergraduate in the lab. Thank Will Porterfield, Will Skyrud, Antonio Del Rio Flores, and Zhijuan Hu for their patient guidance and training during my early years in the lab. Thank Wenlong Cai, Yongle Du, and Kaimin Jia for generously sharing their knowledge, answering my countless questions, and assisting me with troubleshooting experiments. Special thanks Rui Zhai and Yoo-Sung Ko for their invaluable collaboration on the salivabactin projects. Additionally, I would like to express my appreciation to Xuejun Zhu; although we have not yet met, your research materials have greatly facilitated my work on the alkyne projects. Moreover, I am thankful to my mentees, Julie Kim, Breanna Yang, Eloise Liu, Yuxin Xie, Bailey Daymo, and Caroline Luo, for helping conduct experiments and playing a crucial role in supporting me.

I also want to express my profound gratitude for the funding provided by the National Institutes of Health (NIH), the support extended by the California Institute for Quantitative Biosciences (QB3-Berkeley) and the invaluable platform offered by the College of Chemistry at UC Berkeley. In addition, I am appreciative of our collaborators, including the Muthiah Kumaraswami lab, the Ting Xu lab, and the Greg Challis lab, for their generous support of the projects included in this dissertation. Furthermore, I am grateful to the professors on my qualifying exam committee and dissertation committee, including Prof. Michael Marletta, Prof. Michelle Chang, Prof. John Dueber, and Prof. Jonathan Rittle, for their constructive feedback and guidance.

Finally, I wish to extend my heartfelt appreciation to my family, my boyfriend, and my close friends. Deepest thanks to my parents, Haiyin Gu and Hongzhen Cao, for their unconditional love and support. They have always prioritized my needs and well-being above their own, laying the foundation for my success and shaping me into the person I am today. Especially for my mom, I understand that it was a difficult decision for her to allow me to leave our homeland and pursue studies in the US. However, she recognized that this was my aspiration and gave me her blessing. In addition, I would like to thank my boyfriend, Junzhi Wang, for love, support, and unwavering confidence in me. His presence and support have been a constant source of strength and encouragement for me. Last but not least, my sincere thanks go to my close friends for their companionship, understanding and empathy throughout the journey.

The completion of my doctoral degree would never have been possible without the assistance of all the above-mentioned people.

Chapter 1. Introduction

1.1 Natural Product Biosynthesis

Natural products are small molecules produced by diverse natural sources with a long history of significant applications. Nearly 50% of approved new drugs from 1981 to 2019 are natural products or related to natural products, including natural product derivatives and synthetic natural product mimics.[1] The structural complexity and diversity in scaffolds and functional groups of natural products are leveraged for their pharmaceutical potential in treating various public health concerns, especially cancer and infectious diseases.[2–4] Some well-known examples of natural products with clinical applications include erythromycin,[5] rapamycin,[6] paclitaxel[7] and quinine[8] (**Figure 1-1**), which are antibiotic, immunosuppressive, anticancer and antimalarial drugs, respectively. In addition to therapeutic applications, natural products are also widely used in areas such as agriculture, food, dietary supplements, cosmetics and biomaterials.[9–13] Considering the significance of natural products, researchers have been dedicated to revealing the scientific logic behind production and developing engineering strategies to generate products of value or higher yields over the decades.

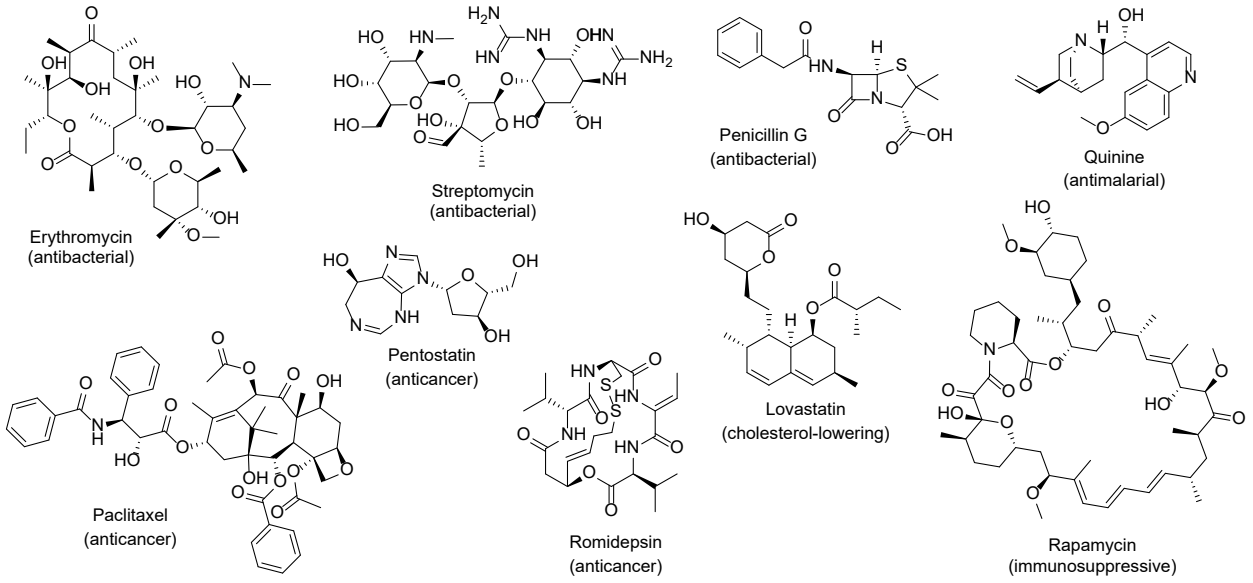


Figure 1-1. Examples of natural products with pharmaceutical applications.

Natural products are secondary metabolites produced by various organisms, such as archaea, bacteria, fungi, plants and animals. The process of generating natural products from simple substrates in living organisms is regarded as biosynthesis. Based on biosynthetic origins, natural products can be categorized into major groups including polyketides (PKs),[14] non-ribosomal peptides (NRPs),[15] ribosomally synthesized and post-translationally modified peptides (RiPPs),[16] terpenes,[17] glycosides,[18] alkaloids,[19] hybrids,[20] and others. Many PKs and NRPs, composing a significant number of natural products with pharmaceutical potential, are produced by modular (type I) polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), which have attracted tremendous attention for mechanistic understanding and pathway engineering. These megasynthases contain multiple enzyme activities in each module and act as an assembly

line to take monomers for modification and polymerization in a thio-templated fashion (**Figure 1-2**).[21] Polyketide chains grow by the iterated addition of malonyl units (or methylmalonyl units), in which the C–C bonds are forged during chain elongation by decarboxylative Claisen condensations. NRPs are formed by the condensation of amino acids, by C–N (amide) rather than C–C bond formation. The PK and NRP chains thus grow as a series of elongating acyl or peptidyl thioesters, in which the sulfur atom of the thiol group is the terminus of a phosphopantetheine (PPant) arm covalently attached to carrier protein domains embedded in assembly lines. Further chemical complexity can be achieved by inclusion of rare building monomers,[22] tailoring enzymes,[23] and hybrid PKS-NRPS pathways.

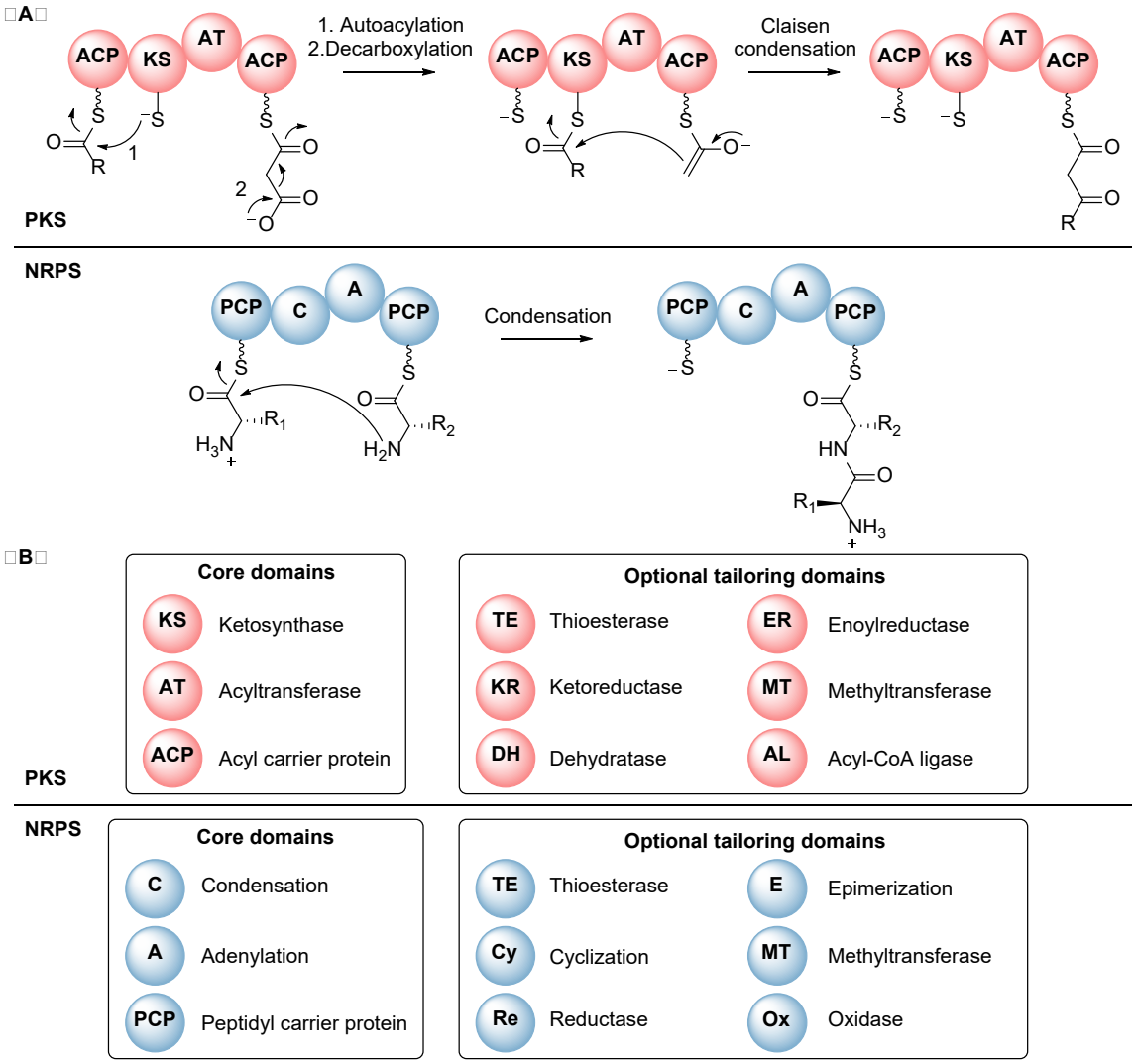


Figure 1-2. Enzymatic machinery of PKSs and NRPSs. **(A)** PKS and NRPS chain elongation mechanisms. **(B)** Core and optional domains in PKS and NRPS. Schematic abbreviations used throughout the dissertation are shown.

The dissertation will limit the discussion of natural products to PKs/NRPs derived from bacteria. Biosynthetic genes in bacteria are typically located in close physical proximity to each other on the genome and are often co-regulated. Thus, these clusters of biosynthetic genes are referred to as biosynthetic gene clusters (BGCs), responsible for the formation of specific natural

products. The work presented in the dissertation contributes to elucidating a biosynthetic pathway of a terminal alkene containing natural product and an unprecedented TE domain. Furthermore, based on the understanding of the biosynthetic pathway, heterologous expression and biosynthetic engineering were employed for enhanced production of the natural product. In addition to terminal alkene containing natural products, investigations were conducted on terminal alkyne formation in natural products and utilization of the terminal alkyne biosynthetic machinery to generate alkyne-labeled natural products.

1.2 Biosynthetic Pathway Engineering

Due to the modular nature and homology of enzymes from diverse pathways, PKSs and NRPSs have been a focal point of engineering efforts to create novel compounds or increase compound titers.[24] Based on the components of building blocks, biosynthetic engineering can be categorized into two classes, starter unit engineering and extender unit engineering. In PKSs, starter unit engineering is more common because extender units are typically well-defined, unless rare acyltransferase (AT) domains are adopted or mutated. The AT domains serve as gatekeeping enzymes in PKSs.[25] For NRPSs, starter units can be various small molecules like fatty acids while extender units are usually different amino acids. The incorporation of amino acids is regulated by the gatekeeping enzymes, adenylation (A) domains.[26] The promiscuity of megasynthases toward their building monomers has also been observed in quite a few PK/NRP biosynthetic pathways, enabling the incorporation of non-native monomers in final products.[27] Therefore, biosynthetic pathway engineering can be classified into four categories based on engineering strategies, which are precursor feeding, site-directed mutagenesis, combinatorial biosynthesis, and engineering of tailoring enzymes (**Figure 1-3**).[28] The examples of these engineering strategies are provided below.

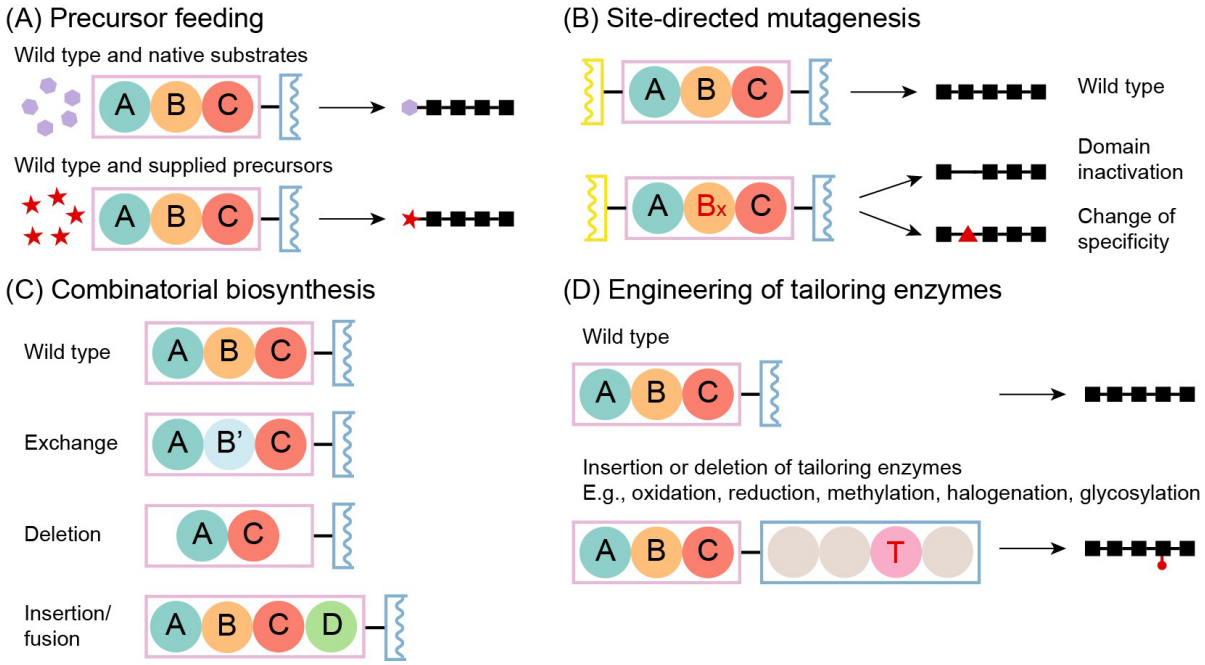


Figure 1-3. Schematic of four biosynthetic pathway engineering strategies. The circles and letters represent domains and rectangles represent modules. Adapted from "Recent Advances in Re-engineering Modular PKS and NRPS Assembly Lines" by C. Beck, *et al*.[28]

Firstly, precursor feeding or precursor-directed biosynthesis (PDB) generates new compounds by providing artificial analogs of native substrates based on the promiscuity of modular enzymes. One of the major challenges of PDB is the competition between native substrates and fed precursors. Thus, if possible, the deletion or suppression of dedicated biosynthetic pathways of native substrates can create a relatively clean background for production of desired PK/NRP analogs containing non-native monomers. Microcystins, the most well-known cyanobacterial toxins, are derivatized by PDB to generate a series of clickable microcystins. By supplementing different azide- or terminal alkyne containing amino acid analogs into the cultivation medium of microcystin-producing cyanobacteria strains, researchers generated more than 40 different clickable microcystins based on the flexibility of substrate acceptance of these strains.[29] Another famous example of PDB is the engineering of rapamycin-type molecules, rapamycin and FK506/FK520. The starter unit of rapamycin is 4,5-dihydroxycyclohexa-1,5-dienecarboxylic acid (DHCHC), derived from the chorismate pathway. Its Production could be abolished by knocking out a single chorismatase gene, *rapK* or *fkbO*, in rapamycin and FK506 pathway, respectively.[30,31] Different starter unit substrates can be fed and accepted by downstream biosynthetic machinery to produce variants of novel analogs with varying titers.[32,33] Besides starter unit engineering, extender units can also be diversified via PDB. FK506 contains a unique allylmalonyl extender unit that is synthesized by a dedicated PKS and associated modification enzymes. Deletion of *tcsB*, a KS encoding gene involved in allyl biogenesis, disrupted this allylmalonyl extender unit biosynthesis. Subsequently, three FK506 analogs were produced upon feeding of trans-2-hexenoic acid, 4-methylpentanoic acid and 4-fluorocrotonic acid.[34,35]

Secondly, site-directed mutagenesis is widely used in biosynthetic engineering due to the rapid development of structural biology, bioinformatics, and computational modelling. It is considered a minimally invasive approach because of its limited effect on the internal structure of the PKS/NRPS systems. In many cases, site-directed mutagenesis is used to alter the substrate specificity of gatekeeping enzymes such as AT domains in PKSs and A domains in NRPSs. The promiscuous AT domain from the antimycin biosynthetic pathway was used as a template to produce derivatives of the larger 12-membered ring depsipeptide JBIR-06 (as opposed to the 9-membered ring antimycin) with alternative PKS extender units.[36] The PKS modules of antimycin and JBIR-06 are quite homologous; however, the antimycin AntD-AT can accept large, diverse substrates while the JBIR-06 SmlC-AT cannot. Inspired by the critical role of "VDYASH" in the AT domain of the erythromycin PKS for substrate recognition, the corresponding active site motif of SmlC-AT, "LRIAPH", was replaced by "MPAAAH" from AntD-AT. This mutated SmlC-AT domain enabled the incorporation of short to medium fatty acyl chains in place of the naturally occurring dimethyl group. These atypical extender units were incorporated to the depsipeptide backbone, and JBIR-06 analogs were produced in practical yields. In addition to expanding substrate specificity of gatekeeping enzymes, site-directed mutagenesis is also used to inactivate or mutate tailoring enzymes. To produce an analog of a quinomycin depsipeptide in *Escherichia coli*, the methyltransferase (MT) domain was inactivated by site-directed mutagenesis to generate non-methylated products.[37] The thioesterase (TE) domain in pikromycin biosynthetic pathway was mutated by only one residue, resulting in the alteration of offloading mechanisms and the production of diastereomeric macrolactones.[38]

Thirdly, combinatorial biosynthesis is a discipline that seeks to combine different natural product biosynthetic pathways via genetic engineering to reprogram PKSs/NRPSs and produce new chemical structures.[28,39] Although it is depicted in **Figure 1-3** at the level of domains, It can actually be implemented at the level of subdomains, domains, modules or subunits. The highly

homologous and seemingly modular nature of PKS/NRPS megasynthases were recognized early on for the potential to swap domains from different pathways to produce novel natural product analogs. Many new analogs have been synthesized following this strategy; however, domain swapping has been and remains a significant challenge. Domain swaps often yield inactive or inefficient enzymes.[40,41] The complexities of intra- and inter-protein interactions are still being elucidated, and ideal swap sites are difficult to identify. Despite these challenges, recent investigations have identified promising sites and methods for incorporating diversity. For example, pyoverdine is a fluorescent NRP siderophore produced by fluorescent pseudomonads. The *Pseudomonas aeruginosa* NRPS PvdD contains two modules, each of which incorporates an L-threonine residue at the C-terminal end of pyoverdine. After engineering attempts, researchers realized that the native PvdD condensation (C) domain exhibited stronger selectivity than A domains for the incorporated amino acid substrate. Hence, substitution of both the C and A domains of PvdD generated high yields of rationally modified pyoverdines in two instances, with these pyoverdines having either a lysine or a serine residue in place of the terminal threonine.[42] Not only can combinatorial biosynthesis generate new compounds, but it can also increase compound titers. Most type I PKSs contain only one acyl carrier protein (ACP). Three ACPs were rationally designed and genetically installed in module 6 of 6-deoxyerythronolide B synthase, which resulted in an approximately 2.5-fold increase in the total amount of PKs produced *in vitro*.[43] Apart from domain engineering, exchange, deletion and incorporation of modules or subunits are also conducted to generate novel products based on high degree of homology among genes of PKS/NRPS systems.

Lastly, engineering tailoring enzymes is a powerful strategy to increase the structural diversity of PK and NRP molecules by altering glycosylation, methylation, oxidation, reduction, or halogenation patterns, which may significantly impact the bioactivities of compounds. Enduracidins and ramoplanin are structurally and functionally closely related lipodepsipeptide antibiotics known for their activity against multidrug resistant Gram-positive pathogens, including MRSA. Each peptide contains one chlorinated non-proteinogenic amino acid residue. In the mutant strain SfΔ30, the enduracidin halogenase gene *orf30* was deleted and complemented with the ramoplanin counterpart *orf20*, leading to the generation of numerous novel analogs with antimicrobial activities.[44]

1.3 Terminal Alkene Containing Natural Products and Biosynthesis

Terminal alkenes can easily undergo chemical derivatization to introduce various desirable functionalities into natural products, such as epoxides, hydroxyl groups and ketones.[45–48] In addition to their innate reactivity, terminal alkene moieties can enhance biological activities of PKs, as exemplified by the improved drug tolerability and efficacy of a synthetic epothilone analog, sagopilone, which exhibits anticancer activity.[49] There are a few known examples of terminal alkene containing PKs/NRPs, including FK506, haliangicin, curacin A, and tautomycetin (**Figure 1-4**).[50–53] These natural products possess potent bioactivities, including immunosuppressive, antifungal, anticancer and antibiotic activities, respectively. Besides PKs/NRPs, terminal alkene containing hydrocarbons derived from bacteria[54] also have tremendous importance and industrial values. Traditionally, 1-alkenes are exclusively sourced from petroleum, which is nonrenewable and energetically costly. The bioproduction of 1-alkenes offers a sustainable alternative with reduced environmental impact. Taking all of these into account, researchers have expressed interest

in terminal alkene containing natural products and studied their biosynthesis over the past few decades.

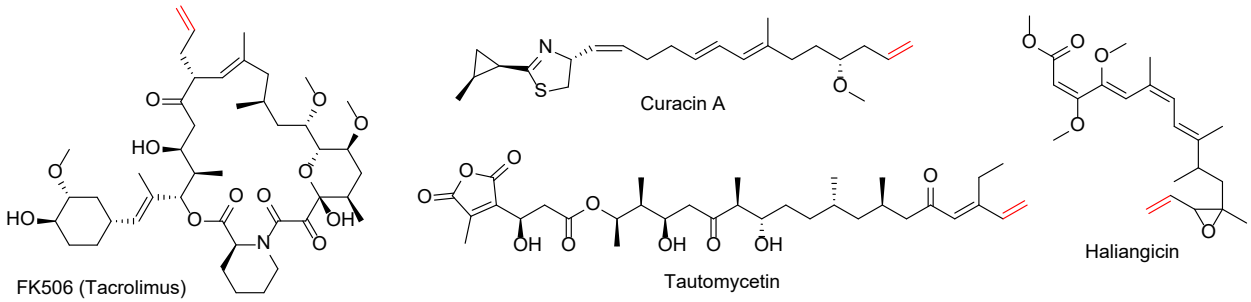
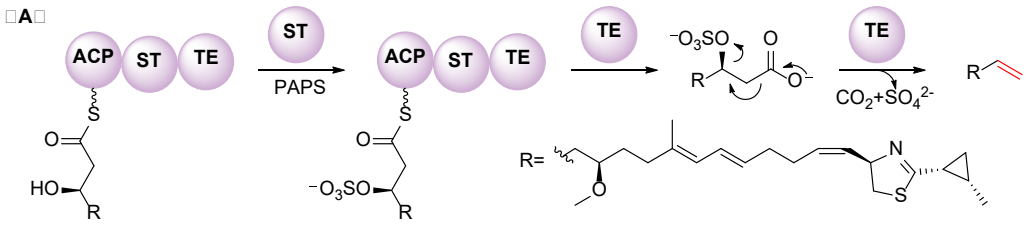


Figure 1-4. Examples of terminal alkene containing natural products.

Different from internal alkenes, which are often generated via the action of ketoreductase (KR) and dehydratase (DH) domains in PKSs, the biosynthetic mechanisms of terminal alkenes in natural products are diverse (**Figure 1-5**). In curacin A, a terminal alkene is formed through a unique offloading mechanism involving a sulfotransferase (ST) and thioesterase (TE) domain.[55,56] The ST domain transforms the *β*-hydroxyl of the penultimate chain elongation intermediate into a *β*-sulfate, an excellent leaving group that is positioned chemically to facilitate decarboxylative elimination in the presence of the terminal carboxylate following TE-mediated hydrolysis of the thioester. The formation of tautomycetin terminal alkene occurs after the chain has been released from the PKS, although the process has yet to be fully characterized.[57,58] Tailoring enzymes in TtnCDFI are proposed to be involved in the terminal alkene formation. Another mechanism of terminal alkene formation occurs through the action of acyl-CoA dehydrogenases (ACAD), which oxidize the *γ,δ*-position of a fatty acyl-CoA or ACP, as observed in the haliangicin and FK506 pathways.[34,59] Specifically, TcsD, responsible for forming the terminal alkene of the allylmalonyl-CoA extender unit in FK506 biosynthesis, acts regioselectively at the *γ,δ*-position and only on *α,β*-unsaturated substrates. The regioselectivity is controlled by a combination of bulky residues in the active site and a lateral shift in the positioning of the flavin adenine dinucleotide (FAD) cofactor within the enzyme.[60] In terms of terminal alkene hydrocarbons, enzymes can perform oxidative decarboxylation through different mechanisms. One example is $OleT_{JE}$, a cytochrome P450 fatty acid peroxygenase/ decarboxylase that catalyzes the conversion of long-chain fatty acids into 1-alkenes in *Jeotgalicoccus* sp. 8456.[61] Furthermore, the Zhang lab reported UndA, a nonheme diiron oxidase for 1-undecene biosynthesis in *Pseudomonas*, and UndB, a membrane-bound desaturase-like enzyme also capable of converting lauric acid to 1-undecene.[54,62]



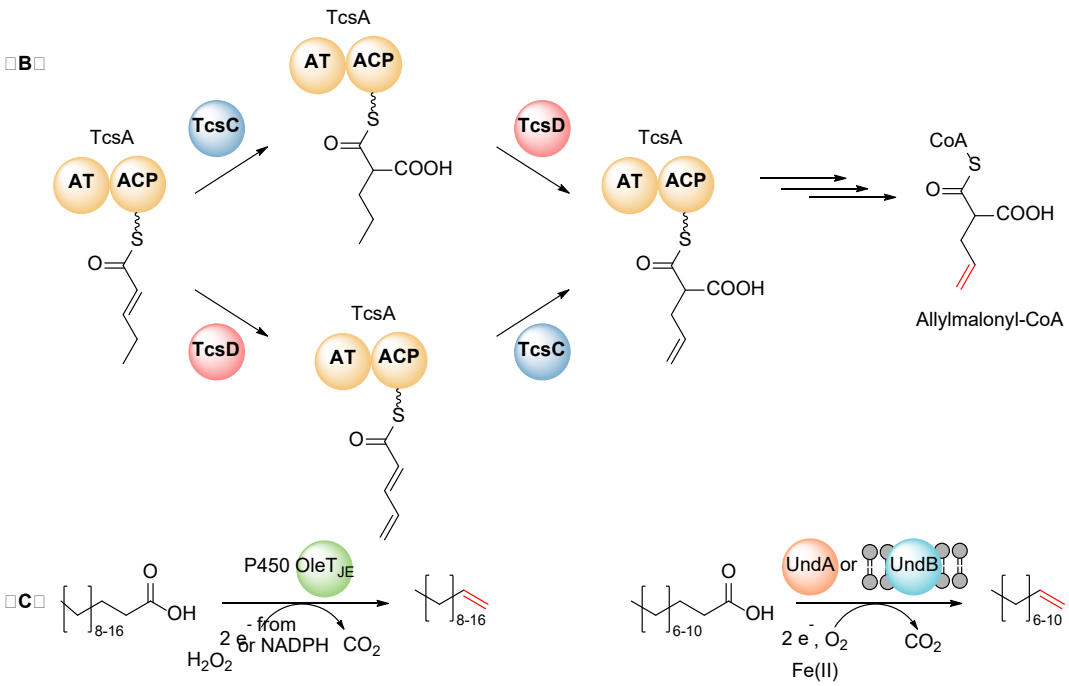


Figure 1-5. Examples of biosynthetic enzymes of terminal alkene formation in natural products. **(A)** Terminal alkene formation in the curacin A biosynthetic pathway. Adenosine 3′-phosphate 5′-phosphosulfate PAPS is the sulfonate group donor. **(B)** Proposed steps of allylmalonyl-CoA biosynthesis. TcsC is crotonyl-CoA reductase while TcsD is ACAD. **(C)** Known enzymatic reactions for 1-alkene biosynthesis.

In Chapter 2 and 3 of this dissertation, we will discuss the elucidation and engineering of the biosynthesis of a recently discovered terminal alkene containing PK/NRP hybrid. We investigated the unique offloading mechanism from the assembly line resulting in terminal alkene formation. Biosynthetic pathway engineering was performed to increase the titer of the terminal alkene containing natural product.

1.4 Terminal Alkyne Containing Natural Products and Biosynthesis

The terminal alkyne is a functionality widely used in organic synthesis, material science, pharmaceutical science and bioorthogonal chemistry.[63–66] Particularly, the terminal alkyne is useful for the alkyne–azide cycloaddition reaction, commonly known as “click” chemistry.[67] Therefore, considering the chemical stability in biological environments, the rarity in biology, and the selectivity in bioorthogonal reactions, the terminal alkyne moiety has become one of the most powerful tools in chemical biology with the applications of discovery, diversification, quantification, visualization and mode-of-action elucidation of biomolecules.[68] Despite the wide applications of terminal alkynes, few terminal alkyne containing natural products have been discovered, and even fewer biosynthetic enzymes for terminal alkyne formation have been characterized so far. Understanding the biosynthetic machinery of terminal alkyne formation facilitates the engineering of biosynthetic pathways to introduce this valuable functionality into various biomolecules. Hence, it is an urgent demand to study terminal alkyne containing natural products and their biosynthesis. It is worth noting that since the polyyne moiety possesses distinct chemical properties and is derived from different biosynthetic mechanisms compared to the terminal alkyne moiety,[69] polyynes will not be discussed here.

Terminal alkyne containing natural products have been identified across various organisms, including bacteria, plants, fungi, marine sponges, and insects (**Figure 1-6**).[70,71] Many of them exhibit potent and broad biological activities, such as antibacterial, antitumor, anti-parasite, anti-malarial, and anti-HIV activities.[72] There are a number of enzymes characterized for acetylenic natural products, including fatty acids, PKs, PK/NRP hybrids, amino acids and meroterpenoids.[70] However, enzymes specifically responsible for terminal alkyne formation are limited to two known enzyme classes, JamB-like nonheme diiron desaturases/acetylenases, and BesB-like pyridoxal 5′-phosphate (PLP)-dependent lyases (**Figure 1-7**).[73,74]

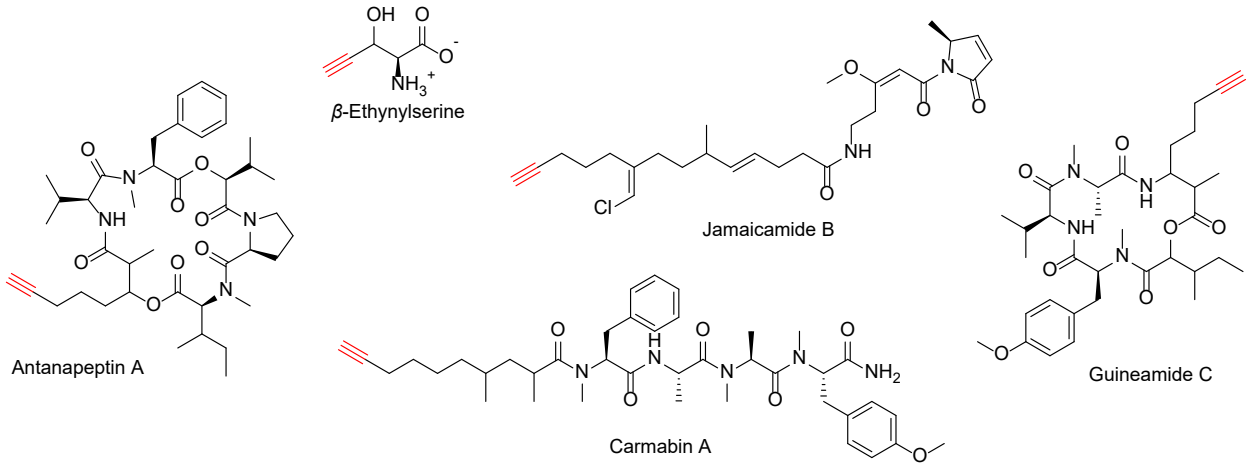
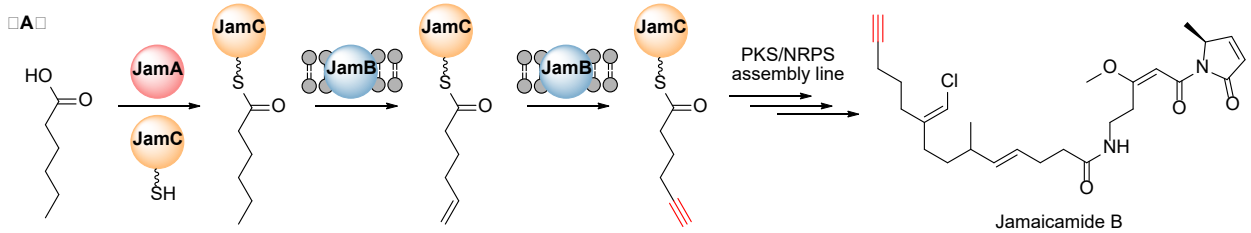


Figure 1-6. Examples of terminal alkyne containing natural products.

JamB-like nonheme diiron desaturases/acetylenases were first reported by the Zhang group in 2015.[73] Acetylenases are a special family of desaturases that catalyze O_2-dependent dehydrogenation of C–C bonds in a diiron dependent mechanism.[75] Inspired by the natural product, jamaicamide B, isolated from the marine cyanobacteria *Moorea producens*, the Zhang group scrutinized the three-gene operon, *jamABC*, and revealed that JamB is the first bifunctional terminal desaturase/acetylenase in a microbial PKS/NRPS pathway to incorporate a terminal alkyne moiety. In this operon, JamA is a homolog of fatty acyl-ACP synthetase and JamC is an ACP. To generate the terminal alkyne, JamA activates and loads hexanoic acid onto JamC, and the resulting hexanoyl-JamC is modified by JamB to yield 5-hexynoyl-JamC as a starter unit for the downstream PKS/NRPS assembly line. Notably, JamB, as a membrane-bound enzyme, is also the first exception to the conventional view that only soluble desaturases work on acyl-ACPs, whereas membrane desaturases use acyl-CoAs or lipids as substrates.[76]



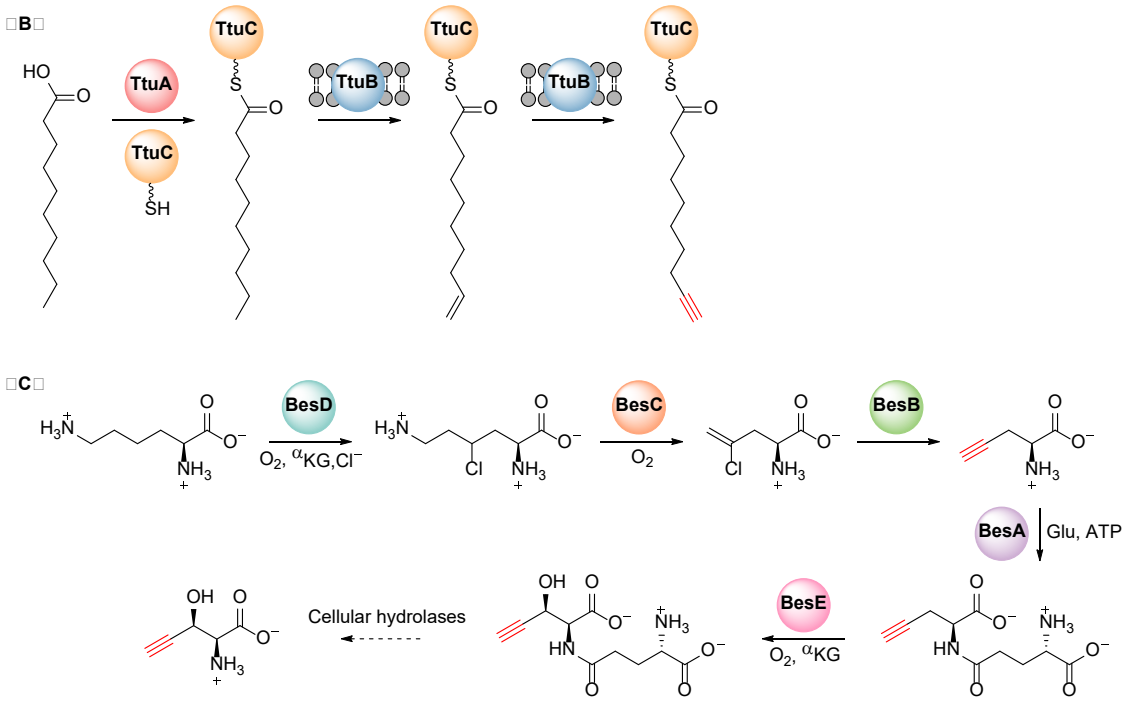


Figure 1-7. Known biosynthetic enzymes for terminal alkyne formation. JamB-like nonheme diiron desaturases/acetylenases, exemplified by **(A)** JamB and **(B)** TtuB. **(C)** BesB-like PLP-dependent lyases.

Genome mining revealed at least 80 gene clusters encoding homologs of JamABC across diverse bacterial genera, providing a great potential to expand this molecular toolbox. The Zhang group later discovered a homologous terminal alkyne biosynthetic pathway comprised of TtuABC from *Teredinibacter turnerae* T7901.[77] The biosynthetic machinery encoded by *ttuABC* closely resembles that of *jamABC*, with the major difference in the chain length of fatty acid substrates.

Apart from the biosynthetic enzymes originating from fatty acid synthase (FAS)/PKS, the Chang group investigated a series of enzymes, BesABCDE, in the biosynthesis of a terminal-alkyne amino acid, *β*-ethynylserine (βes) in *Streptomyces cattleya*.[74] First, the halogenase BesD chlorinates the C*γ* position of L-lysine, followed by an unusual oxidative C–C bond cleavage by the oxidase BesC to yield a terminal alkene. Then, the PLP-dependent enzyme BesB catalyzes *γ*-elimination of chloride and isomerization to form the terminal alkyne. BesA and BesE continue to modify L-propargylglycine to form the final product, βes. Notably, the presence of a *γ*-halogen as a leaving group is required to form the terminal alkyne, which suggests the chlorination happens prior to the acetylenation step. Exploration for BesA-D homologs revealed additional *Streptomyces* spp. with the potential to produce terminal alkyne containing amino acids. Moreover, characterization of the βes biosynthetic pathway paves the way for introducing halogen, terminal-alkene or terminal-alkyne into peptides and proteins, offering a variety of possible downstream applications.

With these enzymes for terminal alkyne formation characterized, it is promising to develop biosynthetic machinery into a portable toolbox to generate alkyne-tagged natural products. precursor-directed biosynthesis (PDB) is a common method to generate labeled natural products.[29] However, the coexistence of diffusible precursors and final products with the same chemical handle introduces substantial background in the PDB production system, making it incompatible with *in situ* bioorthogonal chemical transformations. Thus, d*e novo* biosynthesis of terminal

alkynes as a labeling method is significantly advantageous. Firstly, HsPKS1, a promiscuous type III PKS, was selected to couple with JamABC or TtuABC to generate alkyne tagged PKs,[73,77] which was successful. Due to the tolerance of the fatty acyl chain length and the recognition of the ACP, HsPKS1 is a good candidate for accepting alkynyl starter units. Furthermore, for conventional type I PKSs, engineering strategies such as docking domain adoption and site-directed mutagenesis were used to increase the recognition between JamC/TtuC and downstream PKSs.[78] Further details about terminal alkyne biosynthetic pathway engineering will be provided in Chapter 5.

In Chapter 4, we will discuss more mechanistic questions about JamB/TtuB. Various methods were used to solubilize JamB/TtuB and biochemical studies were conducted to shed light on their functions. In Chapter 5, we will focus on engineering biosynthetic pathways to incorporate terminal alkynes into lipopeptides. Two NRPS model systems were demonstrated to show the feasibility of engineering strategies.

Chapter 2. Biosynthesis of Salivabactin and Characterization of Its Biosynthetic Enzymes

2.1 Introduction

The human body is inhabited by a vast number of microorganisms which form a complex ecological community and influence the human physiology, in the aspect of both health and diseases.[79] For example, *Streptococcus mutans*, a natural and primary inhabitant of the human oral cavity, is known to be a primary driver of ecological succession culminating in dental caries.[80] Moreover, *Streptococcus pyogenes*, otherwise known as group A streptococcus (GAS), is an exclusive human pathogen that causes mild, superficial diseases such as pharyngitis and impetigo as well as life-threatening invasive diseases such as necrotizing fasciitis and streptococcal toxic shock syndrome.[81,82] GAS infections are among the top ten infectious causes of human mortality worldwide.[81] In contrast, *Streptococcus salivarius* K12 (SAL), originally isolated from the saliva of a healthy child,[83] shows bactericidal activity *in vitro* against several pathogens including GAS and thus has been developed into an over-the-counter oral probiotic for years.[84,85] Therefore, it is significant to study oral microbial strains and understand the molecular details of pathogenic or probiotic effects.

Recent studies have revealed that the human microbiota have the potential to produce natural products mediating microbe-microbe and host-microbe interactions.[86,87] Oral microbes are rich sources of PKS/ NRPS BGCs, and more importantly, nearly half of these BGCs correlate with oral health or disease based on cross-referencing against meta-genomes and meta-transcriptomes derived from individuals in healthy or diseased states.[88] Investigating these BGCs and elucidating the structure and function of PKs/NRPs produced by oral microbiota will thus yield valuable biological insights or tools to modulate critical biological processes. As mentioned above, SAL has many known health benefits. However, multiple clinical trials demonstrated inconclusive efficacy of SAL for the prevention and treatment of GAS pharyngitis.[89] Hence, our lab, in collaboration with Kumaraswami lab in Houston Methodist, was motivated to study the molecular mechanisms underlying the role of SAL in oral cavities, particularly the roles of secondary metabolites.[90]

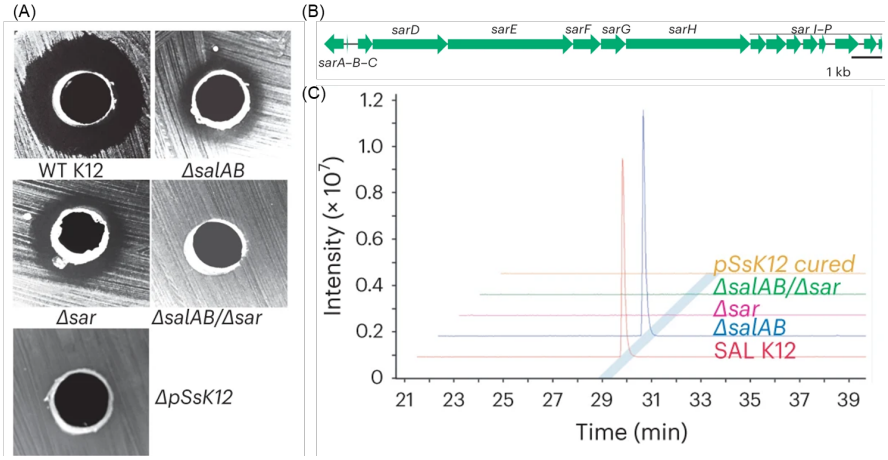
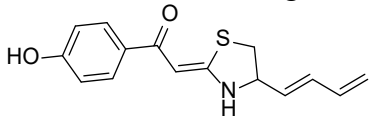


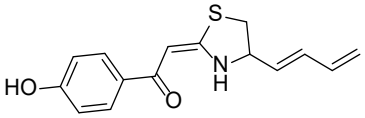
Figure 2-1. Discovery of a new antibiotic from SAL. (A) *In vitro* anti-GAS activities of various SAL strains. The indicated SAL strains were placed in a well on an agar plate containing GAS lawn. The inhibitory activity was

assessed by monitoring the presence or absence of inhibitory zone around SAL growth after 16 h incubation. (B) Organization of the 16-gene operon encoding *sar* BGC. (C) Comparison of LC–MS extracted ion chromatogram traces of the metabolic extracts from SAL and its mutant derivatives, showing the production of salivabactin associated with the *sar* BGC. $EIC^+ = 274.09 \pm 0.01$ (EIC, extracted-ion chromatogram) corresponds to salivabactin.[90]

SAL was thought to exert antimicrobial activity *in vitro* via the production of two lantibiotic peptides, salivaricin A2 and salivaricin B.[91] The BGCs (*salA* and *salB*) for these two lantibiotics reside in a megaplasmid (*pSsK12*) and our deletion of megaplasmid (Δ*pSsK12*) abolished SAL anti-GAS activity. Surprisingly, the Δ*salAB* mutant, in which the production of both lantibiotics was abolished, had attenuated but distinct anti-GAS activity (**Figure 2-1. A**), suggesting that SAL produces lantibiotic-independent antimicrobials and the BGCs for the new antimicrobials are also encoded on megaplasmid *pSsK12*. *In silico* analysis revealed a unique PK/NRP hybrid BGC (named *sar*) that was not found in any other sequenced microbial strain from public databases (**Figure 2-1. B**). Due to the well documented role of many PK/NRPs as antibiotics, our lab hypothesized that the *sar* BGC is responsible for synthesizing an antimicrobial metabolite with a new chemical scaffold. Consistent with this hypothesis, our genetic inactivation of *sar* in the Δ*salAB* mutant (Δ*salAB*/Δ*sar*) resulted in complete loss of anti-GAS activity of SAL *in vitro* (**Figure 2-1. A**). Our initial untargeted metabolomic comparisons of WT SAL and its mutant derivatives revealed a major metabolite that was present only in the culture extracts from *sar*-containing strains (**Figure 2-1. C**). The HRMS analysis of the metabolite, named salivabactin, revealed a molecular formula of $C_{15}H_{15}NO_2S$ (calculated for $C_{15}H_{16}NO_2S^+$: 274.0897; observed: 274.0903). To obtain enough salivabactin for structural elucidation, Zhongrui Li, a previous graduate student, purified ~2.5 mg of compound as a white amorphous powder from a total of ~30 L of SAL cultures. Subsequent 1D and 2D NMR analysis revealed salivabactin to be an approximately equivalent mixture of two geometric isomers with a novel molecular scaffold (**Figure 2-2**). Salivabactin is a terminal alkene containing natural product.



Salivabactin A



Salivabactin B

Figure 2-2. Chemical structure of the two salivabactin isomers, salivabactin A and salivabactin B.

The purified salivabactin displayed potent anti-GAS activity (minimum inhibitory concentration, MIC < 2 μg/mL) and caused complete killing of GAS at 12 h post inoculation at 10X MIC concentration, which is comparable to clinically effective anti-GAS antibiotic, penicillin G (**Figure 2-3. A**). Beyond GAS, salivabactin also showed potency against several clinically relevant gram-positive pathogens including group B streptococcus (GBS), *S. pneumoniae*, *S. mutans*, and *Staphylococcus aureus*, while largely ineffective toward a few tested gram-negative bacteria (**Table 2-1**). Kumaraswami lab also demonstrated that salivabactin had protective effects against GAS in an intramuscular mouse model of infection mimicking human necrotizing fasciitis, indicating the *in vivo* potency of salivabactin to treat GAS infection (**Figure 2-3. B**).

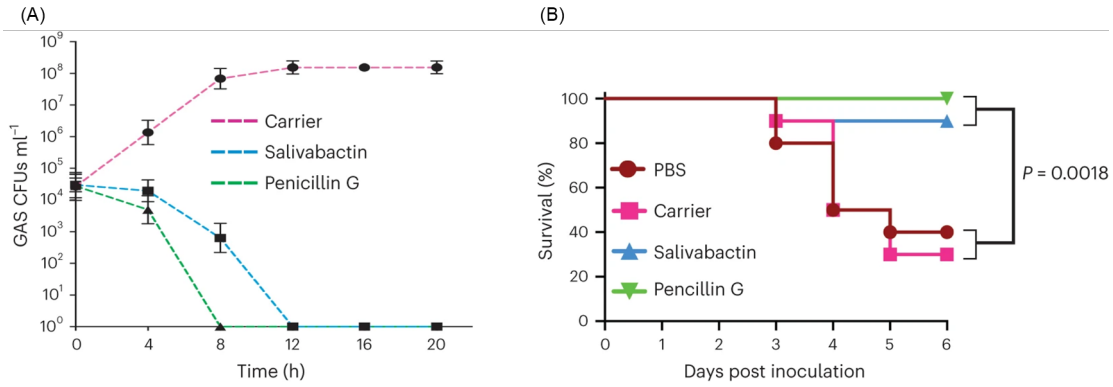


Figure 2-3. Antimicrobial activities of salivabactin. (A) Time-dependent killing assay. Incubation of exponential phase GAS growth with 10× MIC of salivabactin or penicillin G caused complete killing. Data graphed represent mean ± s.d. from three biological replicates. (B) Mice (n = 15 per group) were given 10^7 CFUs of GAS intramuscularly, and antibiotics were given as a single i.m. dose at 1 h post infection. Kaplan–Meier survival curves with *P* values derived by the log-rank test are shown.[90]

Table 2-1. MIC of salivabactin against different bacterial pathogens.

Species and Strain	**Salivabactin MIC (µg/mL)**
Gram-positive pathogens	
Streptococcus pyogenes MGAS10870	2
S. agalactiae ATCC49447	2
S. pneumoniae ATCC49619	2
S. mutans ATCC25175	2
S. dysgalactiae ATCC12394	2
Listeria monocytogenes ATCC19115	2
Staphylococcus aureus ATCC33591	2
Enterococcus faecalis ATCC51299	2
Gram-negative pathogens	
Escherichia coli ATCC25922	>8
Klebsiella pneumoniae ATCC700603	>8
Pseudomonas aeruginosa ATCC27853	>8

After the discovery of salivabactin, the Kumaraswami lab conducted comprehensive investigations into the molecular details of probiotic–pathogenic bacterial interactions *in vivo*. Even though SAL exhibited potent antibiotic activity toward GAS *in vitro*, prophylactic dosing with SAL enhanced GAS colonization in mice and *ex vivo* in human saliva. On co-colonization, GAS responds to a SAL intercellular peptide signal that controls SAL salivabactin production. GAS produces a secreted protease, SpeB, that targets SAL-derived salivaricins and enhances GAS survival. Using this knowledge, they re-engineered probiotic SAL to prevent signal eavesdropping by GAS and potentiate SAL antimicrobials. This engineered probiotic demonstrated superior efficacy in preventing GAS colonization *in vivo*. The findings show that knowledge of interspecies interactions can identify antibiotic- and probiotic-based strategies to combat infection. The impact of those interactions on pathogenesis is critical for patient care and essential to improve probiotic efficacy.

In summary, the discovery of salivabactin establishes the foundation for subsequent research endeavors, including preliminary antimicrobial screening and sophisticated probiotic strain re-engineering. Understanding the activities of *sar* encoded biosynthetic enzymes is essential for designing synthetic biology routes for mass production of useful antibiotics. In addition, *sar* BGC is unique among public strain databases, which indicates a potential of revealing unique biosynthetic pathways and novel enzymes. Therefore, in this chapter, we report the elucidation of salivabactin biosynthesis, *in vitro* reconstitution of the whole biosynthetic pathway and investigations into the releasing mechanism from the assembly line by multidisciplinary expertise in genetics, natural product chemistry, biochemistry, molecular biology, microbiology and structural biology.

2.2 Results

2.2.1 Proposed biosynthetic pathway

Utilizing bioinformatics, the functions of open reading frames encoded by *sar* BGC can be inferred as detailed below (**Table 2-2**). In the discovery paper, stable isotope feeding experiments with [1-$^{13}C_1$] acetate or [1,2-$^{13}C_2$] acetate were conducted to give some biosynthetic insights.[90] Specifically, *sarD-J* are predicted to encode the core *trans* AT PKS/NRPS assembly line to generate the scaffold of salivabactin. The biosynthetic pathway of salivabactin can thus be proposed by leveraging the structural characteristics of salivabactin, preliminary feeding results and bioinformatic analysis (**Figure 2-4**). SarD initiates chain assembly by activating and loading 4-hydroxybenzoic acid (4HBA) followed by one round of PKS extension using malonyl-CoA (Mal-CoA). Upon activation of L-Cys by the A domain of SarE, the heterocyclization (Cy) domain of SarE promotes subsequent condensation and cyclodehydration to generate a five-membered heterocyclic ring. Notably, instead of a typical thiazoline ring, thiazolidine-ene is formed with a mixed configuration, possibly due to a prior unreduced ketide unit.[92,93] The highly discrete machinery of SarE, G, F and H catalyze the next two rounds of PKS extension likely by releasing a β-hydroxyl acid intermediate-335 that undergoes decarboxylation and dehydration to yield salivabactin, although the exact chain releasing mechanism remains obscure. The *trans* AT domain, Sar J, is proposed to interact with megasynthases to fulfill all three rounds of PKS extension. Biochemical experiments are needed for biosynthesis elucidation.

Table 2-2. Deduced roles of open reading frames in the *sar* BGC.

Protein	Locus tag	Deduced role
SarA	*RSSL_00023*	Transcriptional regulator
SarB	N/A	Leaderless peptide
SarC	*RSSL_00024*	Putative DNA-binding protein
SarD	*RSSL_00025*	NRPS-PKS
SarE	*RSSL_00026*	NRPS-PKS
SarF	*RSSL_00027*	Dehydratase
SarG	*RSSL_00028*	Ketoreductase
SarH	*RSSL_00029*	PKS
SarI	*RSSL_00030*	Thioesterase
SarJ	*RSSL_00031*	Acyltransferase
SarK	*RSSL_00032*	4'-phosphopantetheinyltransferase

SarL	*RSSL_00033*	Putative transposase
SarM	*RSSL_00034*	Putative transposase
SarN	*RSSL_00035*	Putative integrase/recombinase
SarO	*RSSL_00036*	Hypothetical protein
SarP	*RSSL_00037*	Hypothetical protein

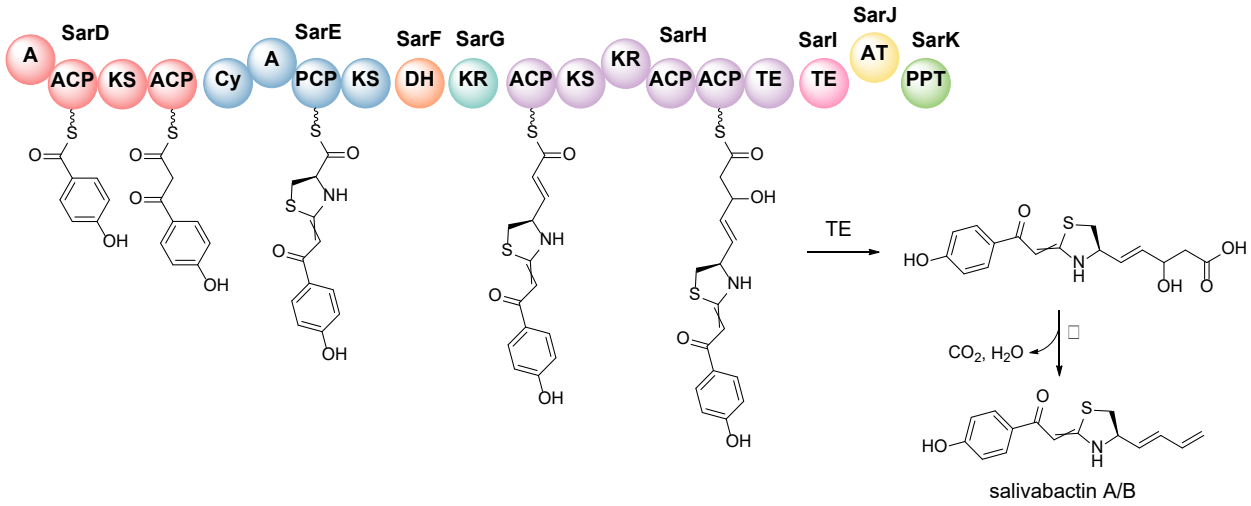


Figure 2-4. Proposed biosynthetic pathway of salivabactin.

2.2.2 Elucidation of salivabactin biosynthesis and *in vitro* reconstitution

To understand the function of each biosynthetic enzyme and define the minimal set of enzymes required for the production of salivabactin, we tried to express and purify all potential biosynthetic enzymes in the pathway with *E. coli* BL21 (DE3)* or BAP1 strain and T7 system, including SarD, SarE, SarF, SarG, SarH, SarI, SarJ and SarK. Among them, SarI is the only enzyme insoluble in *E. coli*. All other enzymes, SarD, SarE, SarF, SarG, SarH, SarJ and SarK, are obtained in their soluble version, verified with sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and saved for further biochemical assays (**Figure 2-5**). The molecular weight of each enzyme from SarD to SarK is 129 kDa, 211 kDa, 50 kDa, 44 kDa, 211 kDa, 34 kDa and 26 kDa, respectively. The estimated yield of each enzyme is 9.9 mg/L, 5.7 mg/L, 28.4 mg/L, 17.3 mg/L, 2.2 mg/L, 3.2 mg/L and 2.5 mg/L, respectively.

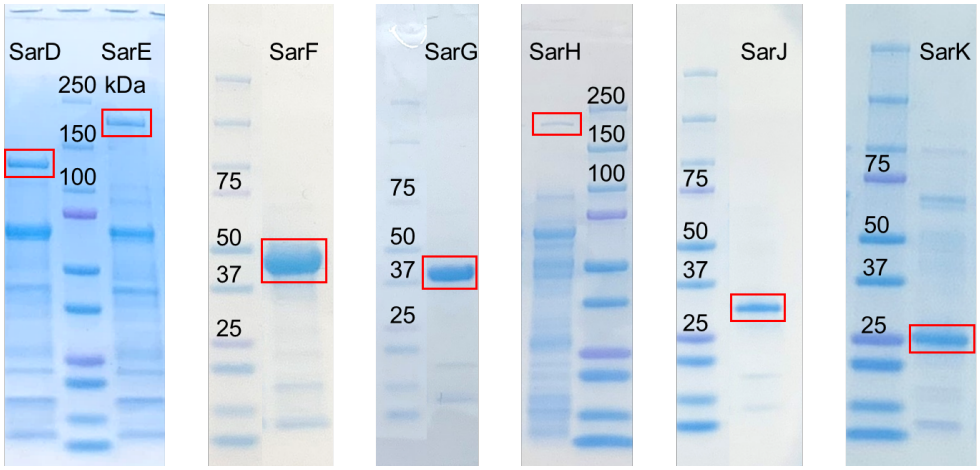


Figure 2-5. SDS-PAGE gel of each biosynthetic enzyme. SarD, SarE and SarH are partially purified.

First, we tested the loading of the starter unit to this assembly line. Since SarD is a megasynthase, it is difficult to directly detect and analyze substrates on ACPs. Instead, we use potassium hydroxide (KOH) to hydrolyze the molecules off from the ACP because after one round of PKS extension, the product is different from the starting material. Small molecules can thus be detected by liquid chromatography/high-resolution mass spectrometry (LC/HRMS) analysis. The proposed substrate of the A domain in SarD is 4HBA. And the extender unit is postulated to be Mal-CoA. Hence, the assay containing SarD, SarJ, 4HBA and Mal-CoA as well as other cofactors was conducted, and assay results were analyzed (**Figure 2-6. A**). We successfully detected the production of 4-hydroxyacetophenone (compound **2**) which is the decarboxylated product of the proposed biosynthetic intermediate, a β-keto carboxylic acid (compound **1**). The commercial standard of compound **2** and negative controls were included for comparison. This coupled assay confirmed the activity of SarD and SarJ. We then continued to examine the activity of SarE by the assay containing SarD, SarE, SarJ. The A domain in SarE is proposed to activate and load cystine to the assembly line, after which the Cy domain is expected to catalyze the formation of five-membered ring. In addition to 4HBA and Mal-CoA, cystine-HCl (Cys) was incorporated for the assessment of SarE activity (**Figure 2-6. B**). As a result, we were able to obtain the production of compound **3** with the support of MS/MS fragmentation pattern (**Appendix A**). Therefore, the step-by-step biosynthesis elucidation of SarD, SarE and SarJ has been confirmed, which is in alignment with the proposed pathway so far.

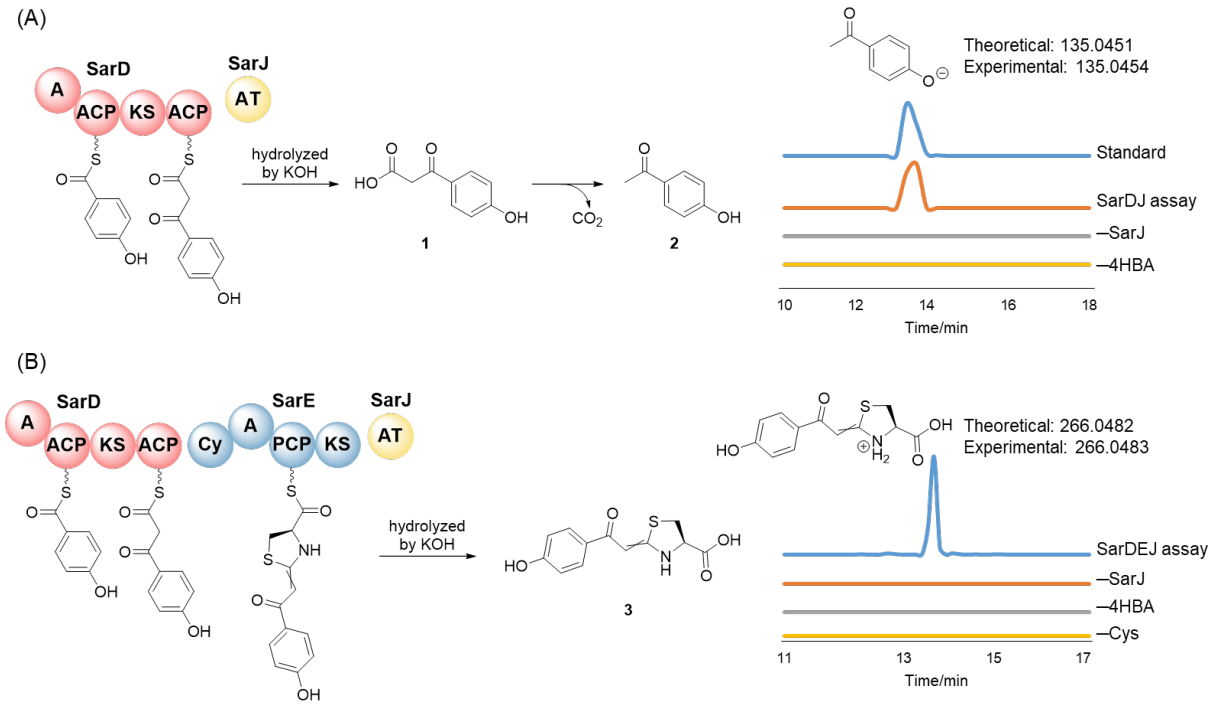


Figure 2-6. Biosynthetic characterization of SarD, SarE and SarJ. **(A)** Biosynthetic pathway involving SarD and SarJ; LC/HRMS extracted ion chromatograms (EIC) of compound **2**. **(B)** Biosynthetic pathway involving SarD, SarE and SarJ; LC/HRMS EIC of compound **3**.

Then, considering the intricacies of the trans-acting machinery involving SarF and SarG, we opted to assess their activities in conjunction with SarH and subsequently identify the corresponding small molecule. Notably, KR domains necessitate NADPH as a cofactor for

functionality.[94] We thus conducted the assay employing enzymes SarD, SarE, SarF, SarG, SarH and SarJ, together with substrates 4HBA and Cys, supplemented with requisite cofactors at room temperature. Delightfully, we were able to detect the production of salivabactin as well as the proposed intermediate-335 (**Figure 2-7**), which demonstrated the success of *in vitro* biosynthesis reconstitution. The MS/MS fragmentation pattern of intermediate-335 is shown in **Appendix A**. Salivabactin is the major product, exhibiting at least 10-fold higher in abundance than the proposed intermediate-335.

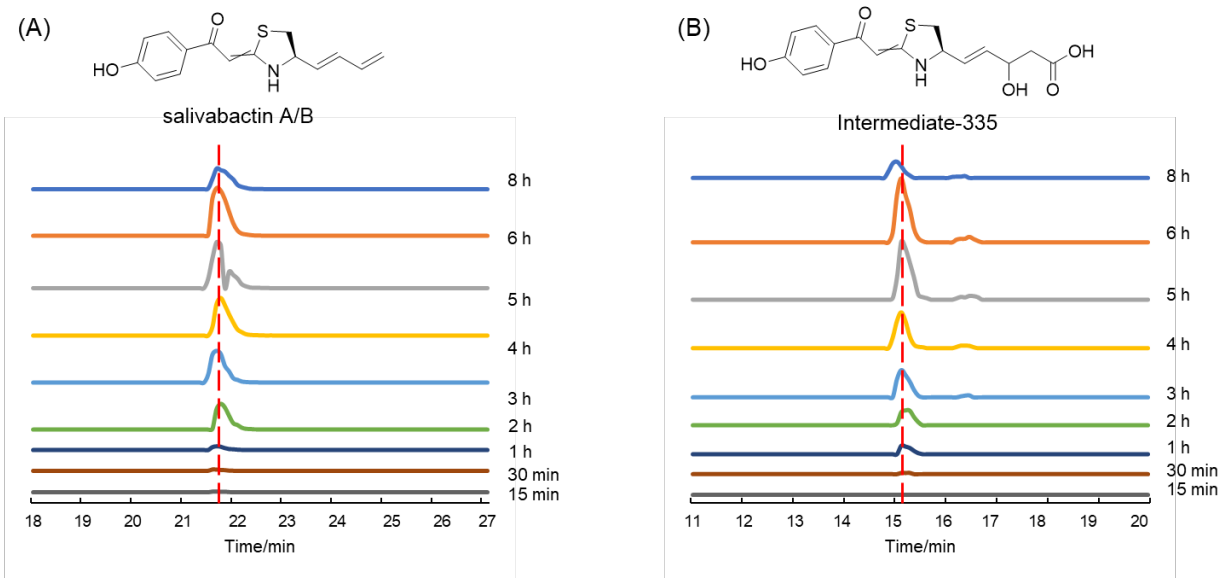


Figure 2-7. Time course results of the production of **(A)** salivabactin and **(B)** intermediate-335. The traces are plotted from LC/HRMS EIC within 10 ppm error. Calculated m/z for salivabactin is 274.0897 and calculated m/z for intermediate-335 is 336.0901. The abundance of salivabactin is 10-fold higher than that of intermediate-335.

Following assay optimization, we tried to identify the minimal set of enzymes required for salivabactin production by systematically excluding each biosynthetic enzyme individually. The absence of SarD, SarE or SarH abolished the production of salivabactin while a significant decrease in production of salivabactin was observed with the omission of SarF, SarG or SarJ (**Figure 2-8**). The absence of SarF or SarG resulted in a 100-fold decrease in abundance of salivabactin while the absence of SarJ led to a 1000-fold decrease. These negative controls indicate the necessity of magasynthases (SarD, SarE, SarH) and trans-AT enzyme (SarJ). Trace amount of salivabactin can be detected with the absence of SarF or SarG, suggesting other enzymes might provide complementary activities or non-enzymatic reactions might take place. For instance, the KR domain in SarH might supplement the function of SarG. However, it is evident that the coordinated action of all six enzymes is requisite for achieving substantial production of salivabactin. In addition, no other enzyme is required for the terminal olefin synthesis. Therefore, the minimal set of biosynthetic enzymes to produce salivabactin includes SarD, SarE, SarF, SarG, SarH and SarJ.

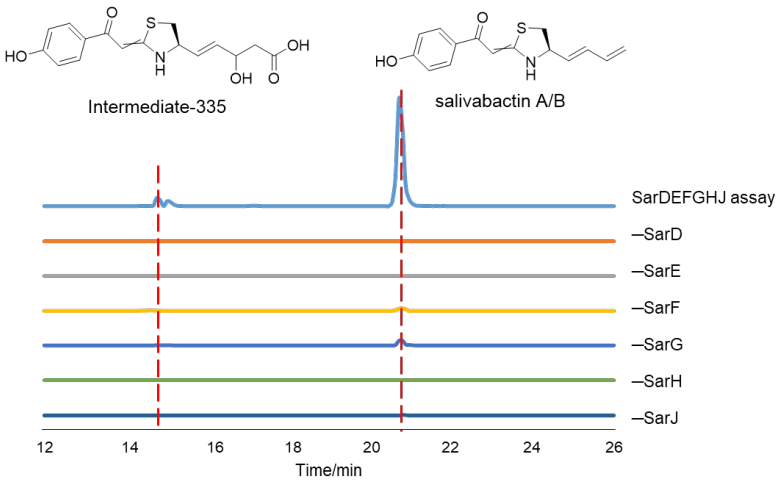


Figure 2-8. *In vitro* reconstitution of salivabactin biosynthesis.

On the other hand, SarI and SarK are not necessarily required for salivabactin biosynthesis *in vitro*. Based on bioinformatic analysis, SarI is a type II TE, which is often used to enhance product yield by correctively removing undesirable substrates or releasing intermediates or final products.[95] The TE domain in SarH adequately facilitates the release of the product from the assembly line, rendering SarI dispensable. Additionally, SarK is a phosphopantetheinyl transferase (PPTase), mediating the transfer and covalent attachment of PPant arms from coenzyme A (CoA) to conserved serine residues of the carrier proteins through phosphoester bonds.[96] Since all the megasynthases were purified from *E. coli* BAP1 strain, which is an engineered strain with a promiscuous PPTase, Sfp, integrated into the chromosome of BL21(DE3).[97] Thus, all the carrier proteins utilized in assays were already equipped with PPant arms, ensuring their activity, the involvement of SarK is superfluous. To summarize, the biosynthetic pathway of salivabactin was elucidated with enzymatic assays and the biosynthesis of salivabactin was successfully reconstituted *in vitro*. The activities of each enzyme in the pathway were validated to correspond precisely with the proposed functions. However, the releasing mechanism underlying the conversion of the potential intermediate-335 into the final product containing terminal alkene remains mysterious, which will be addressed in the following paragraph.

2.2.3 Biochemical studies on TE domain and other biosynthetic enzymes

We hypothesized that the TE domain in SarH serves as the pivotal enzyme facilitating the molecular offloading from the PKS/NRPS assembly line. To test the hypothesis, we generated different SarH/TE constructs and purified all the enzymes from *E. coli* (**Figure 2-9**). We conducted enzymatic assays with different variants of SarH, including wild type (WT) SarH, SarH without TE, the original TE replaced by DEBS TE, supplementation with the original TE and supplementation with DEBS TE (**Figure 2-10. A**). DEBS TE refers to the TE domain from the 6-deoxyerythronolide B synthase (DEBS).[98] It is a canonical TE and can catalyze the macrocyclization at the end of PKS assembly line. As expected, excising TE from SarH eliminated the releasing machinery, leading to significantly diminished production of salivabactin and proposed intermediate-335. Replacing the original TE with DEBS TE yielded a classic hydrolysis reaction with decreased production of salivabactin and increased production of intermediate-335. Adding the original TE back to the SarH without TE recovered the ability of producing salivabactin, suggesting the trans action of TE was still functional. Lastly, reintroducing DEBS TE for trans action failed to restore the production of salivabactin, likely due to inadequate recognition between

the remaining SarH domains and DEBS TE. These biochemical results demonstrated that the formation of salivabactin is catalyzed by the TE domain in SarH. Spontaneous formation of the terminal conjugated olefin is not observed in this instance.

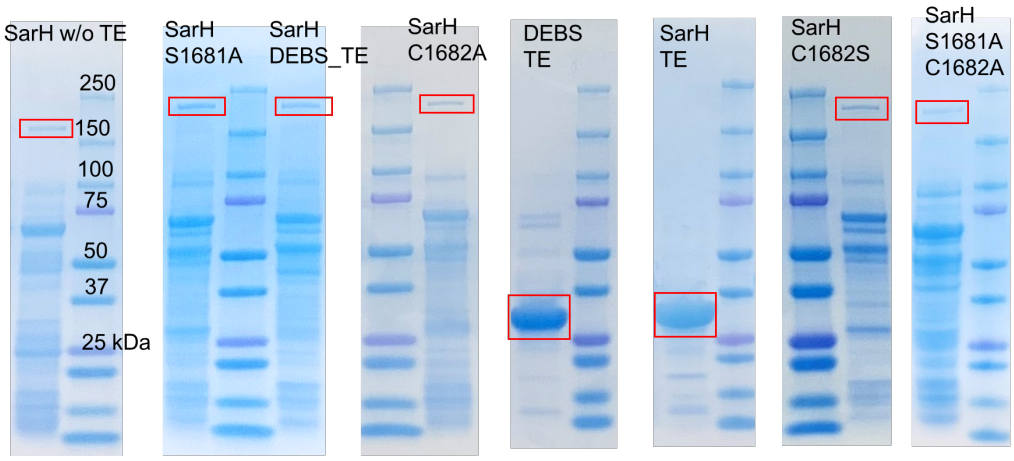


Figure 2-9. SDS-PAGE gel of different versions of TE in SarH.

Given the above enzymatic assay results, it is evident that the TE in SarH serves not only as a canonical TE to produce intermediate-335 with a carboxylic acid, but also possesses unique catalytic capabilities to facilitate decarboxylation and dehydration, resulting in the formation of a terminal alkene functionality. In most cases, TE domains have a serine active site and a conserved Ser-Asp-His catalytic triad. In canonical catalytic mechanism, aspartic acid stabilizes histidine, which serves as a base by accepting a proton from serine. This enables serine to act as a nucleophile, forming a substrate intermediate that is subsequently attacked by water for hydrolysis or by internal functional groups for cyclization.[99] However, According to sequence alignment (**Figure 2-15**), the active site in the TE domain of SarH contains a cysteine residue instead of a serine residue. Additionally, a serine residue is situated immediately preceding the cysteine residue. To comprehensively investigate the releasing mechanism, we employed site-directed mutagenesis and generated several TE mutants in SarH for enzymatic assays (**Figure 2-9**, **Figure 2-10. B**). Mutant C1682A significantly diminished the production of both salivabactin and intermediate-335, indicating the cysteine residue is the active site. Mutant S1681A acted similarly to WT SarH, which means the serine residue is not involved in catalytic activities. Double mutation, S1681A and C1682A, completely abolished the ability to produce either intermediate-335 or salivabactin. Considering canonical TE domains possess Ser as the active site, we also examined mutant C1682S and observed increased production of intermediate-335 as well as significantly decreased production of salivabactin, suggesting mutation from Cys to Ser yields a canonical TE and makes the TE in SarH abolish its unique catalytic capabilities to promote the terminal alkene formation. These results indicated that the unprecedented decarboxylation as well as potentially dehydration catalyzed by TE are probably attributed to the cysteine residue in TE.

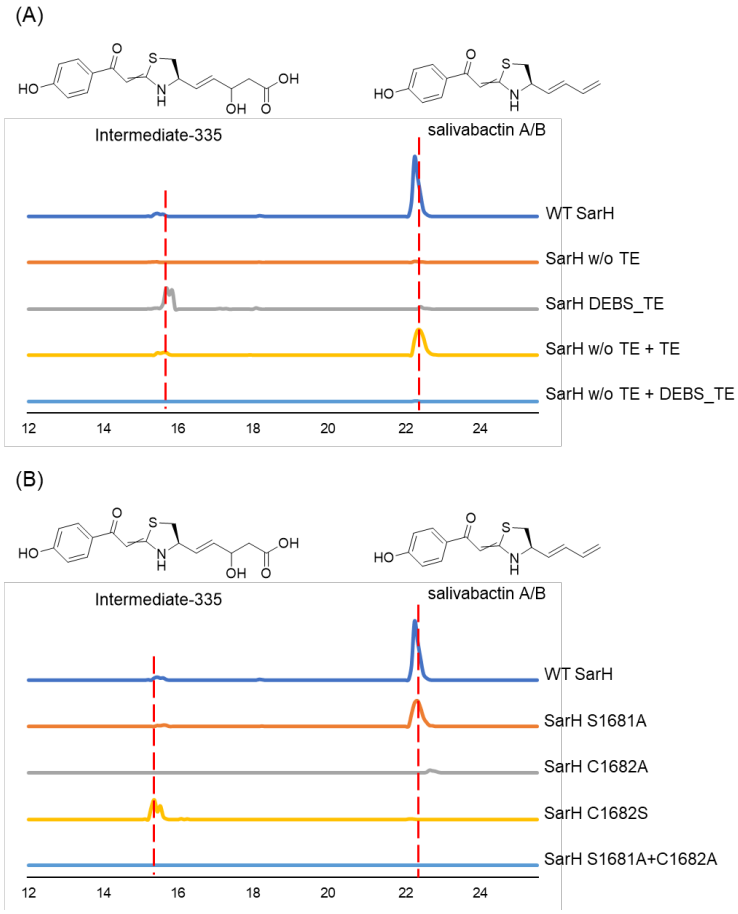


Figure 2-10. Biochemical characterization of different types of TE in SarH. **(A)** The influence of excision, replacement, supplementation of TE on production of intermediate and final product. **(B)** The influence of point mutations of TE on production of intermediate and final product.

All the assays mentioned above are coupled assays involving all requisite biosynthetic enzymes. To directly investigate the catalytic mechanism of the TE domain, we need specific substrates for TE. However, studying the TE domain is challenging due to the complexity of substrates resulting from multiple rounds of PKS/NRPS modification and their connection to the assembly line. Consequently, researchers often seek substrate mimics to overcome this challenge.[100] In our investigation, we tested simple CoA and N-acetylcysteamine thioester (SNAC) substrates, such as octanoyl-SNAC, but these straightforward substrates only verify TE's hydrolysis activity. To scrutinize catalysis involving decarboxylation and dehydration, substrates containing the β-hydroxyl carboxylic acid functionality are imperative. The synthesis of such molecules is currently underway. Meanwhile, we endeavor to purify a substantial quantity of intermediate-335 for both structure characterization and as a substrate for TE assay. Inspired by the enhanced production of intermediate-335 with SarH DEBS TE and SarH C1682S, we expressed different forms of SarH individually together with SarD, SarE, SarF, SarG, SarJ, SarI and SarK in *E. coli* and then extracted the cell culture for compound purification. We tested various growth conditions and determined that the optimal medium is M9 minimal medium (**Appendix B**). Cells harboring biosynthetic genes were inoculated in Luria-Bertani (LB) broth and grew until OD_{600} reached 0.6, after which the cells were transferred to M9 minimal medium for compound production for another 48h. The detailed methodology of heterologous expression will be

elaborated in Chapter 3. As a result, SarH C1682S failed to produce intermediate-335 *in vivo* whereas SarH with DEBS TE succeeded. The majority of intermediate-335 was excreted from cells. Thus, we only need to extract the supernatant of *E. coli* cultures. The purification of intermediate-335 is currently in progress. However, with semi-purified intermediate-335, we performed assays to examine the activity of TE. Unfortunately, TE cannot directly recognize free carboxylic acids as substrates. We therefore synthesized 335-SNAC substrate to emulate the state of molecules tethered to PPant arms. Nevertheless, we only observed a slow hydrolysis reaction without any discernible production of salivabactin, indicating the ACP recognition might play a crucial role in TE function. Hence, we are trying to generate 335-ACP substrates for enzymatic assays, while concurrently exploring the boundary of TE for new TE constructs.

Apart from TE biochemical characterization, we scrutinized activities of other enzymes in the biosynthetic pathway. There are two consecutive ACPs in SarH, which is uncommon in PKS/NRPS hybrid assembly line. We name them ACP2 and ACP3. To determine which ACP is the essential one for activity, we planned to inactivate each ACP individually to evaluate its impact on salivabactin production. Nonetheless, based on sequence alignment of these two ACPs, the residues adjacent to the active site, Ser, are completely identical (**Appendix C**), thereby rendering site-directed mutagenesis impractical for the purpose of inactivating the active site. Therefore, we proceeded to remove each entire ACP from SarH for activity tests (**Figure 2-12**). For both ACPs, it turned out that removing one of them did not affect the activity to produce salivabatcin at all, suggesting that each ACP is capable of accepting upstream substrates and interact with the downstream TE domain. From evolutional perspective, tandem ACPs *in vivo* are likely to increase compound production rate.[43]

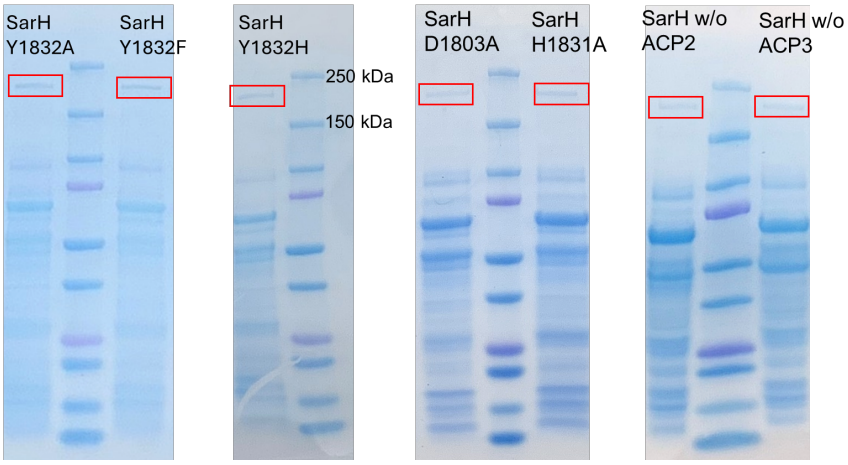


Figure 2-11. SDS-PAGE gel of SarH without ACP and TE mutants in SarH.

2.2.4 Structural biology studies on TE domain and proposed releasing mechanism

In-depth investigation of the TE domain in SarH (SarH_TE) was undertaken through X-ray crystallography studies. In collaboration with Rui Zhai, a postdoctoral researcher in our lab, we solved the crystal structure of the TE domain at a resolution of 1.9 Å. It possesses a typical α/β-hydrolase domain and a lid region composed of two α-helices. (**Figure 2-12. A**) The α/β hydrolase domain consists of residues 1-179 and 224-306, where eight β -strands (β1- β8) are in the central β-plated sheet flanked by helices (αA- αG) on both sides (**Figure 2-12. B**). Residues of the catalytic triad (C137-D258-H286) were confirmed via alignment with the homologous crystal structures obtained from Dali Server[101] (PDB ID: 1AV4,3HI4,1A8U,3HEA) (**Figure 2-13**) and were further

verified via mutagenesis (**Figure 2-11**, **Figure 2-14**). The H286A mutant exhibited a significant reduction in salivabactin production, whereas the D258A mutant retained the ability to produce salivabactin, albeit in a diminished quantity compared to the wild type. This suggests that H286 plays a more critical role in TE activity, while the function of D258 may be complemented by other residues or water.

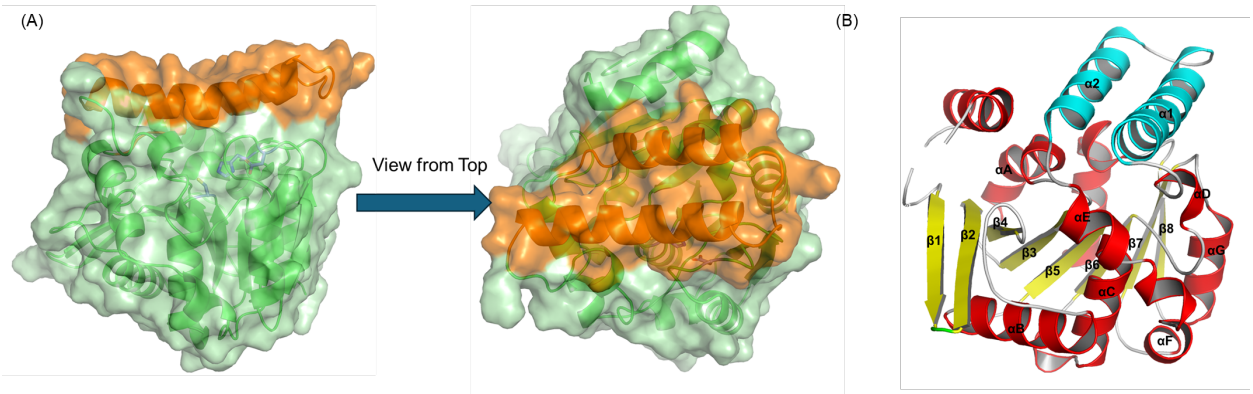


Figure 2-12. Crystal structure of TE in SarH. **(A)** SarH_TE shows a canonical core α/β hydrolase fold (colored green), with a lid region (color orange) composed of two α helices located on top of the core domain. **(B)** Topology structure of TE. α/β hydrolase fold is characteristic of eight β-strands located at the central β-plated sheets which is flanked by α helices on both sides.

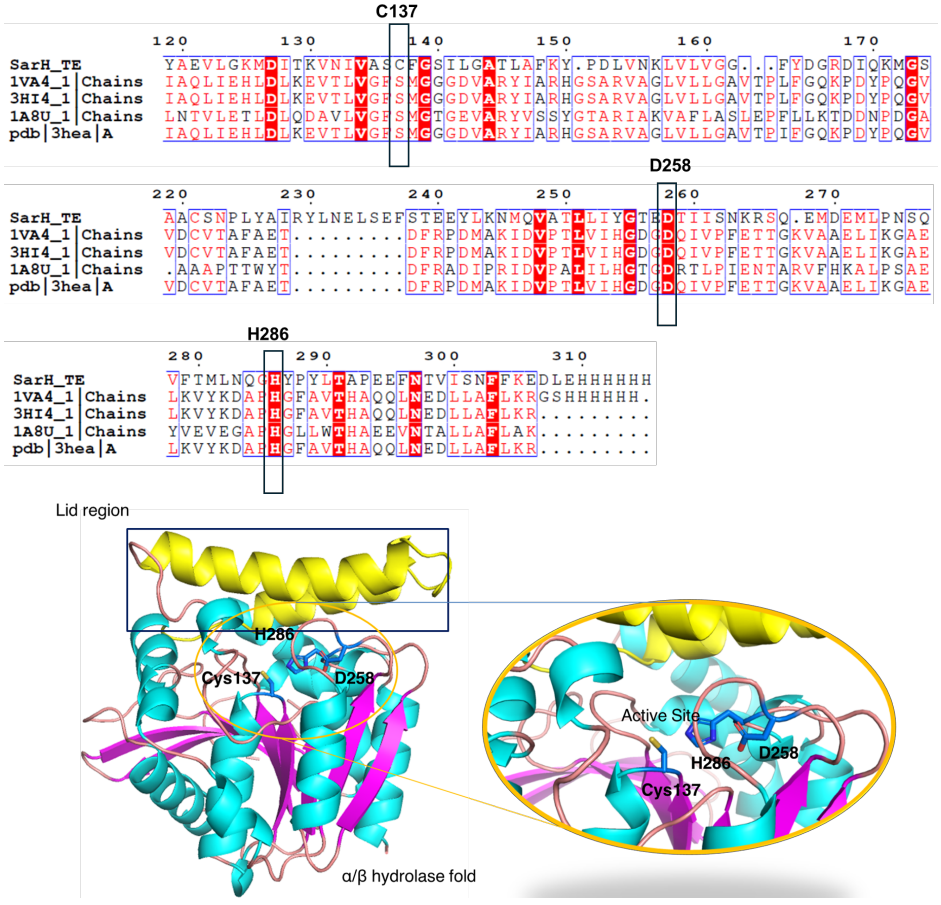


Figure 2-13. Sequence alignment of SarH_TE with structural homologues. Catalytic triad (C137-D258-H286) are conserved and shown in the crystal structure.

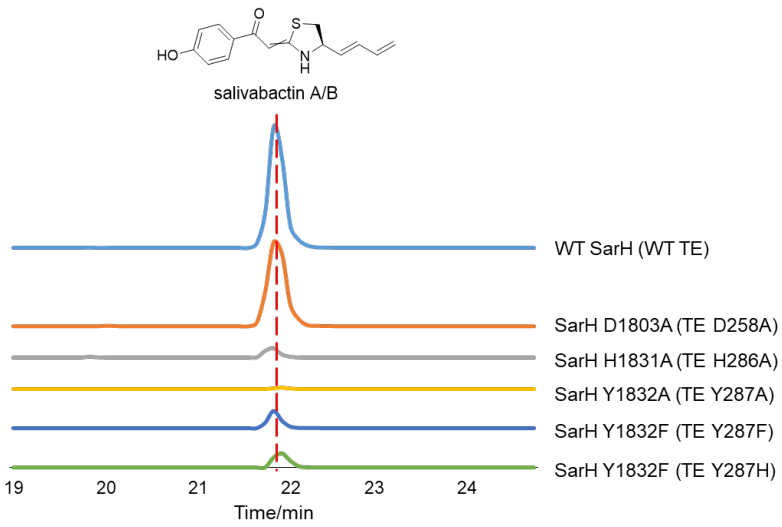


Figure 2-14. Enzymatic assays of TE mutants, measured by LC/HRMS-EIC.

Different from typical TEs in NRPS/PKS pathway, SarH_TE has a noncanonical catalytic nucleophilic residue of cysteine instead of serine.[102] The biochemical assays demonstrates the significance of the Cys residue in the terminal alkene formation. Mutating the Cys residue to a Ser residue converts the TE into a canonical TE, exhibiting only hydrolysis activity. The catalytic triad is located at the top of the core α/β hydrolase domain, oriented towards the cleft between the core domain and the lid, where the substrate can be transferred to TE upon nucleophilic attack by C137 from the carrier protein. Other than the lid region, which typically exhibits greater structural diversity, allowing for varied substrate selection,[103] the biggest difference between SarH_TE and other structural homologues was found to be at the αD (**Appendix D**). In SarH_TE, it was observed that this α helix resides closer to the catalytic center, whereas in its homologue, the longer α helix is positioned further away, resulting in a wide-open space accessible to the catalytic center. Substrates can be bound in the catalytic center through this open space, as shown in 1A8U, while in SarH_TE, substrate binding is blocked by αD (**Appendix D**), suggesting a different binding mode from its structural homologues. Substrates may have access to C137 through a different tunnel (**Appendix D**). To shed light on how the potential intermediate-335 binds inside the tunnel and yields the decarboxylated and dehydrated product, a covalent docking of the substrate linked with C137 was performed in the tunnel (**Figure 2-15**). The model suggests a tight binding mode of 335 within the tunnel, where Y287 is in proximity to the β-hydroxyl group of 335, with a distance of 2.6 Å, thereby facilitating hydrogen bonding interactions. The Mutagenesis study showed that Y287A mutant significantly reduced the production of salivabactin by 100-fold (**Figure 2-14**), confirming the importance of this residue. The mutant Y287F and Y287H decreased salivabactin production by 10-fold compared to WT TE. Moreover, the TE domains in obafluorin and sulfazecin biosynthesis, which also contain cysteine as the active site, were demonstrated to catalyze the formation of a β-lactone and β-lactam in the respective pathway.[104,105] Based on *in vitro* assay results and literature support, a mechanism was proposed as follows (**Figure 2-16**). Activated by H286, Tyr287 acts as a base, deprotonating the hydroxyl group and forming a β-lactone. This β-lactone is unstable and goes though intramolecular decarboxylation to form the terminal olefin. The conjugated double bonds further lower the energy of product, making this step more thermodynamically favorable.

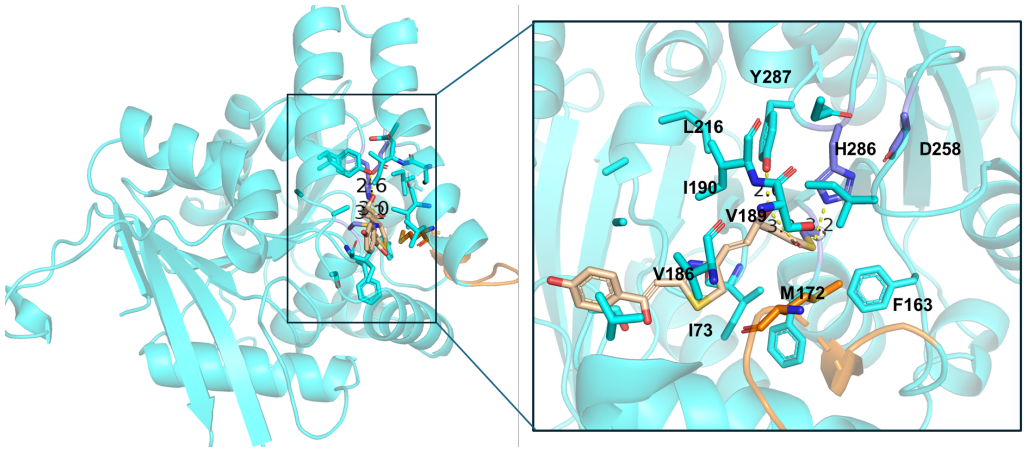


Figure 2-15. Docking model of intermediate-335 covalently linked with C137 in SarH_TE.

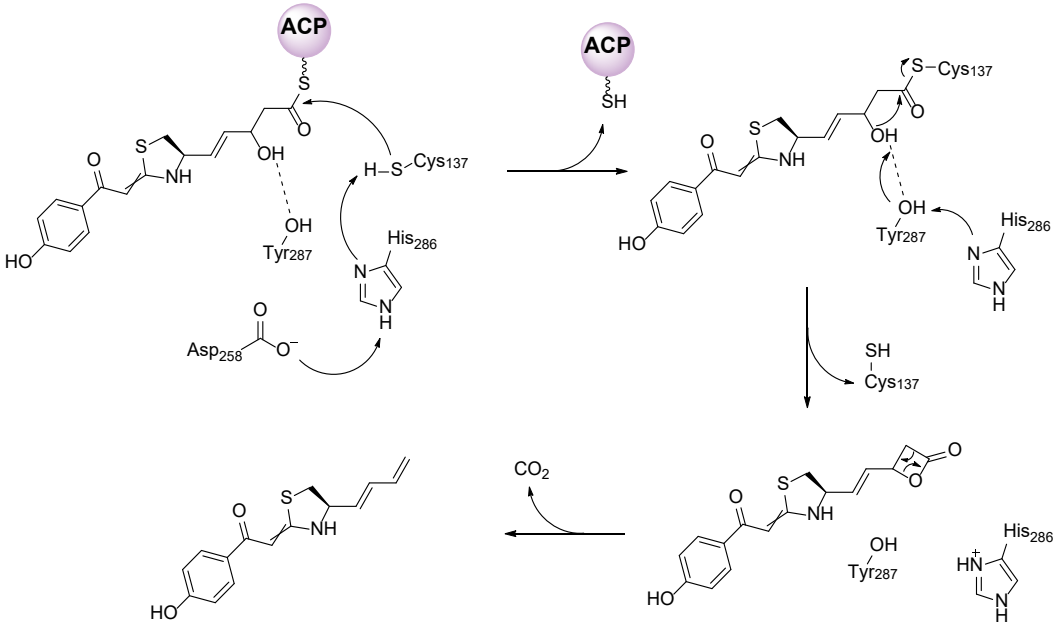


Figure 2-16. Proposed mechanism for the formation of terminal alkene by TE domain in SarH.

2.3 Discussion

The TE domain in SarH constitutes the primary focus of our investigation into the biosynthesis of salivabactin. SarH_TE belongs to a new TE subfamily that is proposed to perform the unique decarboxylation reaction catalyzed by Cys as the nucleophilic residue in the catalytic triad. The phylogenetic tree (**Appendix E**) shows further evolutionary relationship between SarH_TE and other PKS/NRPS TE/TEII (TE16-TE18) based on both sequence similarity and structure similarity.[99] The classification of TE domains was conducted by Benjamin Caswell, *et al*.[99] SarH_TE falls in a distinct clade and shows closer phylogenetic relationships with palmitoyl hydrolase (TE26) represented by YbfF (PDB ID: 3BF7). However, YbfF contains a catalytic tetrad (S89-D113-S306-H234), the orientation and position of which are significantly different from SarH_TE (**Figure 2-17**). The alignment further demonstrates the uniqueness of the TE domain in SarH. It's noteworthy that SphH,[106] a TE domain in the last NRPS module of the biosynthetic pathway of siphonazole, also falls in this clade. Previously SphH was characterized as a regular

TE for hydrolysis to offload the acid intermediate from the NRPS assembly line, and the following decarboxylation/dehydration yielding conjugated terminal double bond happened spontaneously. The determination of the function of SphH relied on a truncated hydroxyhex-4-enoic-SNAC substrate analogue, which might not fully reflect the catalytic mechanism of SphH in the natural product formation. The current crystal structure of SarH_TE and the docking model would provide insights into the new offloading method from NRPS assembly line to form a conjugated double bond, expanding our understanding of TE domain in NRPS/PKS and thus providing guidance for future engineering effort.

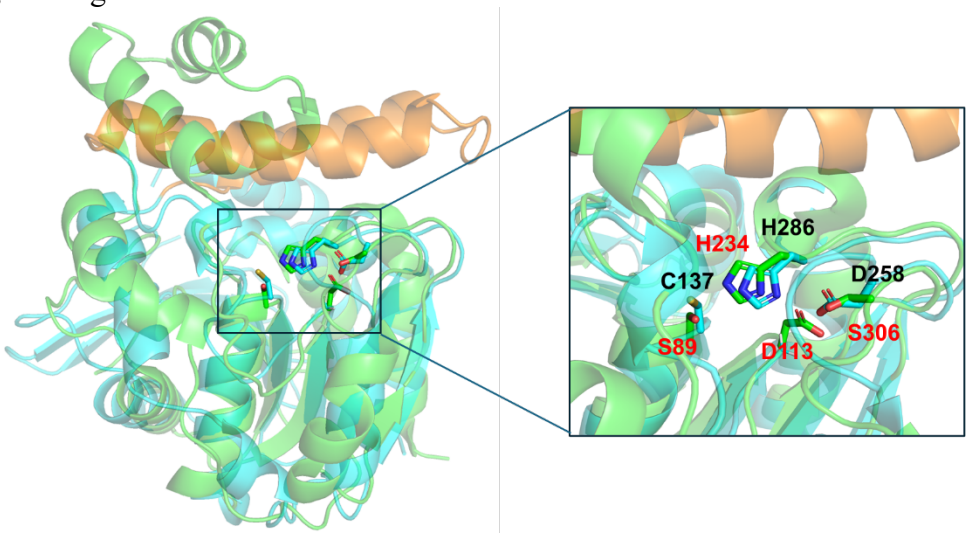


Figure 2-17. Alignment of SarH_TE and YbfF (PDB ID: 3BF7). Other than the difference in the lid region (orange), the catalytic residues are different: Catalytic tetrad (S89-D113-S306-H234) of YbfF are colored in red while catalytic triad (C137-D258-H286) of SarH_TE are colored in black.

In the biosynthetic pathway of salivabactin, we briefly touched upon the function of Cy domain in SarE. Based on structural biology studies, although both C and Cy domains have the same basic fold, their active sites appear distinct. The crystal structure of EpoB in the biosynthesis of epothilones revealed unusual catalytic residues N335 and D449 outside the conserved HHxxxDG motif. The structure also depicted two 20-Å-long channels that create routes for the two tethered substrates to simultaneously reach the buried active site, affording substrate condensation and cyclization. Therefore, the Cy domain performs both the amide bond forming condensation reaction and the cyclization/dehydration to form the five-membered ring structure of thiazoline.[107] However, a thiazolidine-ene ring structure was formed. We hypothesized that this might be owing to a prior unreduced ketide unit. Salivabactin represents only the second known instance of a natural product featuring the thiazolidine-ene ring structure. The first occurrence was observed in forazoline A, which was isolated from marine bacteria, albeit with unidentified BGCs.[93,108] Bioinformatic analysis for both compounds could potentially elucidate the mechanism underlying the formation of this distinctive five-membered ring structure.

Additionally, we explored the influence of tandem ACPs in SarH. Our biochemical characterization demonstrated the equivalent functionality of them. Literature provides several instances of doublet and triplet ACP domains within pathways.[109,110] For example, in the biosynthesis of the antibiotic mupirocin from *Pseudomonas fluorescens*, the removal of each ACP individually results in a basic but diminished pathway activity, underscoring the role of tandem ACP systems in enhancing antibiotic production. Deleting one of the tandem ACPs in SarH could potentially reduce salivabactin production *in vivo*. From the perspective of biosynthetic

engineering, the insertion of multiple consecutive ACPs in pathways may boost the production of natural products.[43]

In summary, we reconstituted the biosynthesis *in vitro* and identified the minimal set of biosynthetic enzymes for salivabactin production. Notably, the identification of a potential intermediate prompted a detailed investigation into the TE domain in SarH. Our comprehensive biochemical characterization of the TE domain confirmed its unprecedented catalytic function with the strategies of site-directed mutagenesis, domain removal, and domain swapping. Furthermore, the crystal structure of the TE domain revealed more insights into the releasing mechanism and the terminal olefin formation. In addition, we investigated the role of tandem ACPs in SarH. The elucidation of the biosynthesis and characterization of biosynthetic enzymes not only advance our understandings of thioesterase activity and PKS/NRPS hybrid pathway, but also provide guidelines for synthetic biology routes for mass production of new antibiotics that overcome the emerging multi-drug resistance.

2.4 Materials and Methods

Chemicals and reagents. Chemicals including 4-hydroxybenzoic acid, 3-hydroxy-benzoic acid and 2-hydroxybenzoic acid were purchased from TCI Chemicals. Malonyl CoA was purchased from CoALA Biosciences. Difco 5x M9 minimal salts were purchased from BD Biosciences. Other chemicals were purchased from Fisher Scientific or Sigma-Aldrich. Phusion High-Fidelity PCR Master Mix (NEB) was used for PCR reactions. Restriction enzymes were purchased from Thermo Scientific. NEBuilder HiFi DNA Assembly Master Mix was used for DNA construct cloning.

Table 2-3. Oligonucleotides used in this study.

Primers	Sequence (5' → 3')	Notes
PDG09_F	GGTTGGGAATTGCAAATGGGAGAATTTTATTCAAAAATAAGAGAAA	Overexpression of SarD
PDG09_R	GGAGATGGGAAGTCATTATCACTCATTTTCTACTCCTTTAAACAAG	
PDG10_F	GGTTGGGAATTGCAAATGAGTGAAAAATATCCACTTTCGG	Overexpression of SarE
PDG10_R	GGAGATGGGAAGTCATTATCATTCATTTACTATTCCTCCTAATACAAC	
PDG22_F	GGTTGGGAATTGCAAATGAATGATAATACAGGTTACATCATACCA	Overexpression of SarF
PDG22_R	GGAGATGGGAAGTCATTAAATTTTTTTTTTTTTGCCACTAAAACCG	
PDG23_F	GGTTGGGAATTGCAAATGACTTTTAATTATCAGATTGGCTCATTT	Overexpression of SarG
PDG23_R	GGAGATGGGAAGTCATTATAATTCATCTAAAACAACCTTTCCTTTTG	
PDG49_F	GAAGGAGATATACATATGGAAAAAACAGTTTTTCT	Overexpression of SarJ
PDG49_R	GTGGTGGTGCTCGAGGTTGTTACCTTCAGCTCC	

PDG50_F	AGAAGGAGATATACATATGCACCACCACCACC ACCACAATTATAATGATGTTAAAAAAAT	Overexpression of SarH
PDG50_R	GGTGGTGGTGCTCGAGATCCTCCTTAAAAAAA TTACTTA	
PDG62_F	CTTTAAGAAGGAGATATACCATGGGAGAATTTT ATTCAAAAATAAG	Co-expression of SarDE and SarFG
PDG66_R	AGCATTATGCGGCCGCAAGCTTATTCATTTACT ATTCCTCCTAATAC	
PDG67_F	TATAAGAAGGAGATATACATATGAATGATAATA CAGGTTACATC	
PDG67_R	GTTTCTTTACCAGACTCGAGTTATAATTCATCTA AAACAACCTTTC	
PDG65_F	GAAGGAGATATACATATGTTAGATAGGATTAGT GAAG	Overexpression of TE in SarH
PDG64_R	GTGGTGGTGCTCGAGATCCTCCTTAAAAAAAT TACTTA	
PDG69_F	CTTTAAGAAGGAGATATACATATGAAAAACCTC ACATTAAAAAT	Overexpression of SarK
PDG69_R	GGTGGTGGTGGTGCTCGAGCAATAATTCAGAT AAGCACTTAAC	
PDG50_F	AGAAGGAGATATACATATGCACCACCACCACC ACCACAATTATAATGATGTTAAAAAAAT	Overexpression of SarH w/o TE
PDG71_R	GGTGGTGGTGGTGCTCGAGTTTAATCGCTAGA TTATTATCC	
PDG72_F	GTCAACATAGTTGCTGCTTGTTTCGGATC	Overexpression of SarH S1681A
PDG72_R	GATCCGAAACAAGCAGCAACTATGTTGAC	
PDG76_F	GTCAACATAGTTGCTTCTGCTTTCGGATCAATT TTG	Overexpression of SarH C1682A
PDG76_R	CAAAATTGATCCGAAAGCAGAAGCAACTATGT TGAC	
PDG50_F	AGAAGGAGATATACATATGCACCACCACCACC ACCACAATTATAATGATGTTAAAAAAAT	Overexpression of SarH w/ DEBS-TE
PDG74_R	ACCGGAGTCCAGCTGTTTAATCGCTAGATTATT ATCC	
PDG75_F	AATCTAGCGATTAAACAGCTGGACTCCGGTAC GCC	
PDG75_R	TGGTGGTGGTGCTCGAGGCTGTTGCCGCCACC CAG	
PDG79_F	AAGAAGGAGATATACATATGCAGCTGGACTCC GGTACGC	Overexpression of DEBS-TE
PDG75_R	TGGTGGTGGTGCTCGAGGCTGTTGCCGCCACC CAG	
PDG68_F	TAAGGAGATATACCATGGGCAATTATAATGATG TTAAAAAAATTATTTA	

PDG84_R	ATTATGCGGCCGCAAGCTTAGCTGTTGCCGCC ACCCAG	Co-expression of SarH w/ DEBS-TE and SarIJK
PDG70_F	TATAAGAAGGAGATATACATATGCAGAGTGAG AATAATATTA	
PDG70_R	GTTTCTTTACCAGACTCGAGTCACAATAATTCA GATAAGC	
PDG85_F	GTTGCTTCTTCTTTCGGATCAATTTTGGGTGC	Overexpression of SarH C1682S
PDG85_R	GATCCGAAAGAAGAAGCAACTATGTTGACTTT TG	
PDG68_F	TAAGGAGATATACCATGGGCAATTATAATGATG TTAAAAAAATTATTTA	Co-expression of SarH C1682S and SarIJK
PDG68_R	ATTATGCGGCCGCAAGCTTAATCCTCCTTAAAA AAATTACTTATT	
PDG87_2_F	CAAAAGTCAACATAGTTGCTGCTGCTTTCGGA TC	Overexpression of SarH S1681A+C1682A
PDG87_R	CAAAATTGATCCGAAAGCAGCAGCAACTATG	
PDG94_F	CAAGGACACGCTCCATATTTAACAGCTCCTGA AG	Overexpression of SarH Y1832A
PDG94_R	GGAGCGTGTCCTTGATTCAGCATAGTGAATACT TG	
PDG95_F	CAAGGACACTTTCCATATTTAACAGCTCCTGAA G	Overexpression of SarH Y1832F
PDG95_R	GGAAAGTGTCCTTGATTCAGCATAGTGAATACT TG	
PDG96_F	CAAGGACACCATCCATATTTAACAGCTCCTGA AG	Overexpression of SarH Y1832H
PDG96_R	GGATGGTGTCCTTGATTCAGCATAGTGAATACT TG	
PDG115_F	CAGAAGCTACAATTATTTCTAATAAACGTTCTC AAG	Overexpression of SarH D1803A
PDG115_R	GAAATAATTGTAGCTTCTGTTCCATAAATAAGA AG	
PDG116_F	GAATCAAGGAGCCTATCCATATTTAACAGCTCC TG	Overexpression of SarH H1831A
PDG116_R	GGATAGGCTCCTTGATTCAGCATAGTGAATACT TG	
PDG117_F	GAATGTCTCTTCAATCACTCAAGAGCTTCAGC ATC	Overexpression of SarH w/o ACP2
PDG117_R	CTCTTGAGTGATTGAAGAGACATTCTTAACTAG TTTCTG	
PDG118_F	CAAGCATATTAATTTCTCAACAGCTTTCTCCAT C	Overexpression of SarH w/o ACP3
PDG118_R	CTGTTGAGAAATTAATATGCTTGATCCATCCCC	

Table 2-4. Plasmids used in this study.

Plasmid	Vector	Description
pDG03	pLATE52	Overexpression of SarD
pDG04	pLATE52	Overexpression of SarE
pDG08	pLATE52	Overexpression of SarF
pDG09	pLATE52	Overexpression of SarG
pDG26	pET24b	Overexpression of SarJ
pDG27	pET24b	Overexpression of SarH
pDG36	pETDuet	Co-expression of SarDE and SarFG
pDG38	pET24b	Overexpression of TE in SarH
pDG41	pET24b	Overexpression of SarK
pDG44	pET24b	Overexpression of SarH w/o TE
pDG46	pET24b	Overexpression of SarH S1681A
pDG47	pET24b	Overexpression of SarH C1682A
pDG48	pET24b	Overexpression of SarH w/ DEBS-TE
pDG49	pET24b	Overexpression of DEBS-TE
pDG52	pCDFDuet	Co-expression of SarH w/ DEBS-TE and SarIJK
pDG53	pET24b	Overexpression of SarH C1682S
pDG56	pCDFDuet	Co-expression of SarH C1682S and SarIJK
pDG58	pET24b	Overexpression of SarH S1681A+C1682A
pDG62	pET24b	Overexpression of SarH Y1832A
pDG63	pET24b	Overexpression of SarH Y1832F
pDG64	pET24b	Overexpression of SarH Y1832H
pDG81	pET24b	Overexpression of SarH D1803A
pDG82	pET24b	Overexpression of SarH H1831A
pDG83	pET24b	Overexpression of SarH w/o ACP2
pDG84	pET24b	Overexpression of SarH w/o ACP3

Table 2-5. Bacterial strains used in this study.

Strain	Description
S. salivarius K12	DNA templates for *sar* genes cloning
E. coli XL1 Blue	DNA cloning and amplification
E. coli BL21 (DE3)*	Recombinant protein overexpression
E.coli BAP1	Recombinant protein overexpression and compound heterologous expression

Construction of plasmids for protein expression in *E. coli*. Individual genes were PCR amplified from the megaplasmid extracted from *S. salivarius* K12. All primers are listed in **Table 2-3**. Gibson assembly was used to ligate PCR products to linear pET24b vector. Thermo Scientific aLICator Ligation Independent Cloning and Expression System was used for plasmid ligation. Plasmids were isolated using a Zyppy Miniprep Kit (Zymo Research) and confirmed by DNA sequencing (UC Berkeley DNA Sequencing Facility).

Overexpression and purification of recombinant protein. Expression strains were grown in

1 L of LB supplemented with 50 μg/mL of kanamycin or 100 μg/mL of ampicillin at 37 °C and 220 rpm until an OD_{600} of 0.6. Cultures were then cooled on ice for 10 minutes before induction with 0.2 mM (isopropyl β-D-1-thiogalactopyranoside) IPTG. Induction of gene expression lasted for 16 h at 16 °C and 200 rpm. The cells were then harvested by centrifugation (6000 x g, 15 min, 4 °C), and supernatant was removed. The pellet was resuspended in 30 mL of lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM imidazole) and lysed with a Q700 sonicator. The insoluble fraction was removed by centrifugation (15,000 rpm, 1 h, 4 °C), and the supernatant was filtered with a 0.45 μM filter before batch binding. Ni-NTA resin (Qiagen) was added to the filtrate at 2 mL/L of cell culture, and samples were allowed to nutate for 1 hour at 4 °C. The protein resin mixture was added to a gravity filter column. The flow through was discarded, the column was then washed with 20-30 mL of wash buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 20 mM imidazole) and tagged protein was eluted in approximately 8 mL of elution buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 250 mM imidazole). Complete elution was determined by Bradford assay. Purified proteins were then concentrated and exchanged into appropriate buffer (50 mM HEPES, pH 7.5, 100 mM NaCl) using Amicon ultra filter units. After two rounds of exchange and concentration, pure protein was removed, and glycerol was added to a final concentration of 8%. Proteins were flash frozen in liquid nitrogen and stored at −80 °C or used immediately for in vitro assays. The presence and purity of enzymes was assessed using SDS-page and concentration was determined using a NanoDrop UV-Vis spectrophotometer (Thermo Scientific).
Molecular weight and yield of purified proteins: pDG38 TE (36 kDa): 11.6mg/L; pDG62 (209 kDa): 1.2 mg/L; pDG63 (209 kDa): 1.3 mg/L; pDG64 (209 kDa): 1.2 mg/L; pDG81 (210 kDa): 1.8 mg/L; pDG82 (210 kDa): 2.0 mg/L; pDG83 (203 kDa): 2.3 mg/L; pDG84 (203 kDa): 2.1 mg/L

Enzymatic reactions involving the activity of SarDJ. A reaction mixture in 50 μL contained 50 mM HEPES (pH 8.0), 2.5 mM $MgCl_2$, 5 mM ATP, 1mM TCEP, 2.5 mM 4HBA, 1 mM Malonyl CoA, and 25 μM SarD. The reaction was initiated by the addition of 5 μM SarJ. The reaction was quenched with the addition of 100 μL cold MeCN. Precipitated proteins were removed by centrifugation. LC/HRMS analysis was performed with 30 μL injection on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2–98% CH_3CN (v/v) with 0.1% (v/v) formic acid over 30 min in H_2O with 0.1% (v/v) formic acid at a flow rate of 0.5 ml/min was used. LC-HRMS/MS analysis was performed using the same program described above with untargeted mass/mass and a collision energy of 10 eV, 15 eV and 30 eV.

Enzymatic reactions involving the activity of SarDEJ. A reaction mixture in 50 μL contained 50 mM HEPES (pH 8.0), 2.5 mM $MgCl_2$, 5 mM ATP, 1mM TCEP, 2 mM 4HBA, 2 mM Cys-HCl, 1 mM Malonyl CoA, 20 μM SarD and 20 μM SarE. The reaction was initiated by the addition of 5 μM SarJ. The reaction was quenched with the addition of 100 μL cold MeCN. Precipitated proteins were removed by centrifugation. The supernatant was subjected to LC/HRMS analysis.

Enzymatic reactions involving the activity of SarDEFGHJ. A reaction mixture in 50 μL contained 50 mM HEPES (pH 8.0), 2.5 mM $MgCl_2$, 5 mM ATP, 1mM TCEP, 2 mM 4HBA, 2 mM Cys-HCl, 2 mM NADPH, 1 mM Malonyl CoA, 20 μM SarD, 20 μM SarE, 20 μM SarF, 20 μM SarG and 20 μM SarH or SarH variants. The reaction was initiated by the addition of 10 μM SarJ. The reaction was quenched with the addition of 100 μL cold MeCN. Precipitated proteins were removed by centrifugation. The supernatant was subjected to LC/HRMS analysis.

Enzymatic reactions involving the activity of TE in SarH. A reaction mixture in 50 µL contained 50 mM HEPES (pH 8.0) and 1 mM (estimated) of 335 or 335-SNAC. The reaction was initiated by the addition of 50 µM TE. The reaction was quenched with the addition of 100 µL cold MeCN. Precipitated proteins were removed by centrifugation. The supernatant was subjected to LC/HRMS analysis.

SNAC substrates synthesis. Roughly 1 mg semi-purified 335 was dissolved in dry CH_2Cl_2 (50 µL) under N_2 and cooled to 0 °C. N-Acetylcysteamine (2.5 µL, roughly 4 eq), DMAP (0.02 mg, 0.05 eq) and Dicyclohexylcarbodiimide (DCC) (0.8 mg, 1.2 eq, dissolved in 50 µL dry CH_2Cl_2) were added successively. The mixture was stirred for 15min at 0 °C, restored to room temperature and stirred overnight. CH_2Cl_2 was removed under vacuum. The residues were redissolved in acetonitrile and centrifuged at 15,000 g for 2 min to remove the precipitation. 335-SNAC was purified by HPLC using an isocratic method of 30% acetonitrile/water. 335-SNAC was confirmed by LC/HRMS and MS/MS fragmentation. HRMS (ESI): calculated for $C_{20}H_{25}N_2O_5S_2^+$: 437.1199, observed: 437.1198.

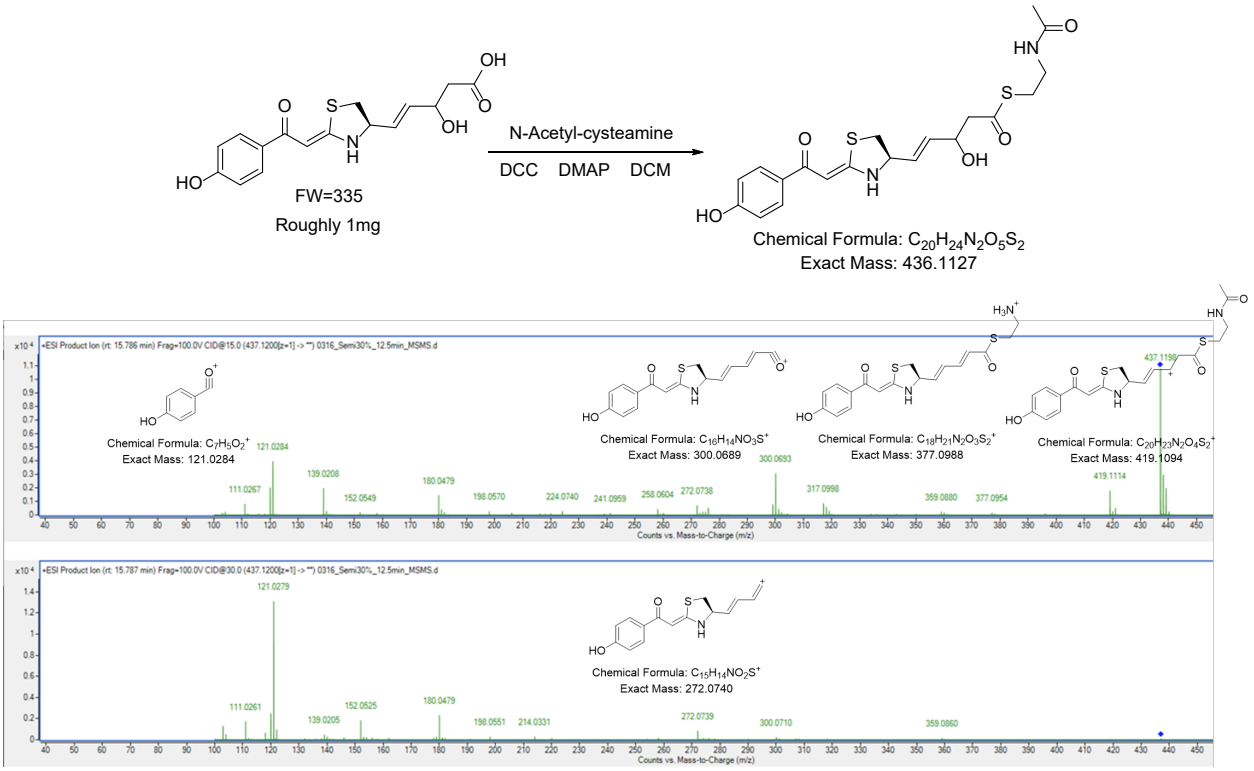


Figure 2-18. Synthesis of 335-SNAC, which was validated with HRMS and MS/MS fragmentation.

Heterologous expression for the production of compound 335. The E. coli BAP1 strain, containing pDG36 and pDG52, were grown in 30 mL LB medium for seeding overnight at 37 °C with 100 µg/ml carbenicillin and 50 µg/ml spectinomycin. The seed cultures were inoculated into 3 L LB medium to grow until OD_{600} reached 0.6. Cultures were then cooled on ice for 10 minutes before induction and substrate feeding. LB medium was removed by centrifugation and cells were resuspended in 1 L F1, MR, M9 minimal medium. M9 medium was found to be the optimal medium for **335** production. Upon transferred to minimal medium, the cultures were added with

0.5 mM IPTG and 0.1 g/L 4BHA. After 48 h of growth at 18 °C, the culture broth and the cell pellets were separated by centrifugation. Compounds were extracted from cell-free supernatant (1L) and cell pellets using ethyl acetate. The solvent was removed by rotary evaporation. The residue was redissolved in methanol and analyzed by LC/HRMS (1 mL culture equivalent injection). LC/HRMS analysis was performed on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2–98% CH3CN (v/v) over 30 min in H2O with 0.1% (v/v) formic acid at a flow rate of 0.5 ml/min was used. LC-HRMS/MS analysis was performed using the same program described above with untargeted mass/mass and a collision energy of 15 eV.

F1 medium recipe. 1L of F1 medium contains 3 g KH_2PO_4, 6.62 g K_2HPO_4, 4 g $(NH_4)_2SO_4$, 150.5 mg $MgSO_4$, 5–20 g glucose, 1.25 mL trace metal solution (optional), 100 μM $Fe(NH_4)_2(SO_4)_2$, and 10 mL 100x vitamin solution (optional). Trace metal solution is composed of 40 mg $ZnCl_2$, 200 mg $FeCl_3 \cdot 6H_2O$, 10 mg $CuCl_2 \cdot 2H_2O$, 10 mg $MnCl_2 \cdot 4H_2O$, 10 mg $Na_2B_4O_7 \cdot 10H_2O$, 10 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ in 1 L of DI water. 100x vitamin solution contains 100 mg/L biotin, 100 mg/L choline chloride, 100 mg/L D-calcium pantothenate, 100 mg/L folic acid, 100 mg/L nicotinamide, 100 mg/L pyridoxal hydrochloride, 10 mg/L riboflavin, 100 mg/L thiamine hydrochloride, 200 mg/L i-inositol, and 8.5 g/L NaCl.[68]

MR medium recipe. 1L of MR medium (pH 6.8) contains 6.67 g/L of KH_2PO_4, 4 g/L of $(NH_4)_2HPO_4$, 0.8 g/L of citric acid, and 5 ml/L of trace metal solution was supplemented with 0.8 g/L of $MgSO_4 \cdot 7H_2O$. The trace metal solution contains 0.5 M HCl, 10 g $FeSO_4 \cdot 7H_2O$, 2 g $CaCl_2$, 2.2 g $ZnSO_4 \cdot 7H_2O$, 0.5 g $MnSO_4 \cdot 4H_2O$, 1 g $CuSO_4 \cdot 5H_2O$, 0.1 g $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, and 0.02 g $Na_2B_4O_7 \cdot 10H_2O$ per liter of solution.

M9 medium recipe. 1L of M9 medium contains 200 mL Difco 5x M9 minimal salts, 20 mL 500 g/L glucose stock solution, 10 mL 80g/L $MgSO_4 \cdot 7H_2O$ stock solution, 1.25 mL trace metal solution. The trace metal solution used here is the same as that in F1 medium.

Chapter 3. Biosynthetic Pathway Engineering of Salivabactin

3.1 Introduction

Since salivabactin has shown potent activity against a few gram-positive human pathogens (**Table 2-1**), it is conceivable that it may play a broad ecological role within the oral community, which necessitates further mode-of-action research. It is essential to find a strategy to produce salivabactin in a reliable and scalable manner. Our collaborator sought help from total synthesis groups to synthesize salivabactin, but their efforts failed, underscoring the importance of biosynthesis for this compound. From the original producing host, SAL, the titer of salivabactin remains relatively low with only 2.5 mg of compound obtained from 30 L of SAL cultures. Additionally, the media utilized for SAL growth, brain heart infusion (BHI) broth or Todd-Hewitt broth plus yeast extract (THY) broth, are notably rich, posing challenges for purification. More importantly, salivabactin is transiently produced by the native host (**Figure 3-1**), creating significant obstacles in accumulating a substantial quantity of the compound. In SAL, the titer of salivabactin starts decreasing after 6-hour growth, probably due to the self-defense mechanism against the potent antibiotic. Therefore, it is imperative to develop new approaches to produce a large quantity of salivabactin stably and make purification processes easier.

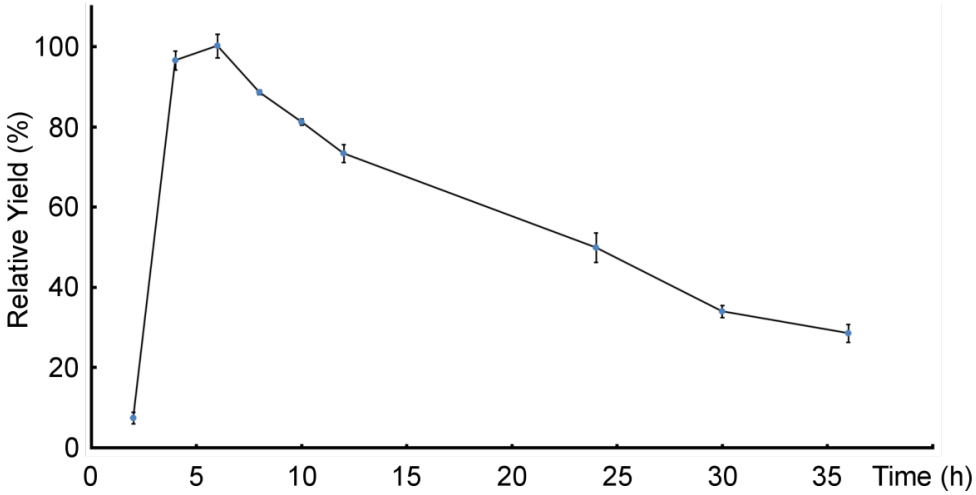


Figure 3-1. Time course for salivabactin production, measured by LC/HRMS, in wild-type SAL strain grown in THY broth at 37 °C. Data are shown as mean ± s.d. (n = 3).

The biosynthesis elucidation of salivabactin A/B facilitates the biosynthetic pathway engineering. First of all, in light of the limitations associated with utilizing the native host, we aim to identify an alternative host for salivabactin production by heterologous expression. Then engineering strategies such as precursor feeding, site-directed mutagenesis and combinatorial biosynthesis can be employed to increase the titer of salivabactin further. Heterologous expression empowers not only the discovery of novel bioactive natural products, but also facilitates the robust production of valuable biomolecules in more amenable heterologous hosts. This methodology also fosters the generation of novel analogs coupled with biosynthetic engineering.[111] Rapid advancements in genome sequencing and genetic engineering techniques has increased the

feasibility of cloning and reconstruction of BGCs in heterologous hosts. Gene-controlling elements must be carefully considered to ensure the successful heterologous expression of BGCs, including promoters, regulatory factors, BGC copy number and their genomic positioning. The heterologous production of natural products may be influenced by specific regulatory and biosynthetic elements within the host. Thus, employing a diverse range of heterologous hosts may ensure more robust heterologous expression.

A notable example of heterologous expression is regarding the biosynthesis of a polyketide, 6-deoxyerythronolide B (DEB), yielding a scalable production route to the erythromycin precursor and establishing a model system that is still used to decipher the intricacies of PKS biochemistry. The erythromycin producing gene cluster was first discovered and annotated in *Saccharopolyspora erythraea* in 1991.[112] Through a combination of structural characterization of the polyketide products, gene sequencing, and chromosomal gene knockout, researchers showed that three dimodule megasynthase PKSs were responsible for synthesizing the DEB backbone of the potent antibiotic erythromycin, named as 6-deoxyerythronolide B synthase (DEBS). In total, the assembly line contains 28 distinct active sites and is greater than 10 kb in size. Enabling the heterologous expression of this gene cluster in a more genetically accessible strain with enhanced fermentation capabilities was an attractive approach to boost compound titer and access new structural derivatives of erythromycin for drug production and discovery. This goal was ultimately achieved through the successful heterologous expression of the DEB pathway in *E. coli*. The first major hurdle was that *E. coli* does not possess a PPTase gene. In polyketide biosynthesis, the PPTase is responsible for activating ACP of the assembly line by covalently tethering a CoA derived PPant arm to a conserved serine residue on the ACP. Absence of a PPTase renders the DEBS modules inactive. The commonly used PPTase gene, *sfp*, was thus inserted into the *E. coli* BL21 (DE3) genome to yield the new strain, *E. coli* BAP1.[97] BAP1 is commonly used today for PKS heterologous expression and recombinant protein expression. The next hurdle was the fact that *E. coli* lacks the capability to produce *(2S)*-methylmalonyl-CoA, the building block for DEB biosynthesis. The *(2S)*-methylmalonyl-CoA biosynthetic machinery from *Streptomyces coelicolor* was thus introduced to the *E. coli* BAP1 strain and verified to be able to produce the desired *(2S)*-methylmalonyl-CoA stereoisomer.[97] The plasmid containing the DEBS machinery was then introduced to this *E. coli* BAP1 chassis strain. Upon fermentation the strain was fed exogenous propionate, and the production of the desired DEB polyketide was verified. This demonstrated that large PKS modules could be successfully expressed in *E. coli*. It also provided a strain that could be scaled to economically produce the erythromycin precursor, DEB. This is an example of a highly complex and synthetically inaccessible scaffold that can be biosynthesized and then semi-synthetically converted into erythromycin or novel erythromycin derivatives. Salivabactin, being a compound that is challenging to synthesize, may also hold the potential for production through heterologous expression.

In this chapter, we report the heterologous expression of *sar* BGC in *E. coli* for salivabactin production and biosynthetic pathway engineering to generate salivabactin analogs *in vivo*. The titer of salivabactin has been increased by 7-fold compared with the native host production through heterologous expression and growth optimization.

3.2 Results

3.2.1 Heterologous expression of *sar* BGC in *E. coli*

E. coli was selected to be the heterologous host for the following reasons. First, according to the *in vitro* reconstitution of biosynthetic pathway, all biosynthetic enzymes except SarI can be successfully expressed and soluble in *E coli*. The essential enzymes required for salivabactin production are purified from *E. coli* and validated for activity through enzymatic assays, indicating that *E. coli* has the capability to process active biosynthetic enzymes encoded by *sar* BGC. Selecting *E. coli* as the heterologous host presents a high probability of yielding the final product, salivabactin. Second, *E. coli* is fast growing and also the most thoroughly characterized bacterial host, features an impressive range of genetic manipulation protocols. Both biosynthetic pathway engineering and strain engineering can be feasibly accomplished in *E. coli* at a later stage for the purpose of compound titer enhancement.

The essential biosynthetic enzymes encoded by *sar* BGC were acquired via recombinant protein expression in *E. coli*. For the enzymes containing ACPs or peptidyl carrier proteins (PCPs), SarD, SarE and SarH, *E. coli* BAP1 strain was used for the installation of PPant arms onto carrier proteins. For other enzymes, SarF, SarG, SarI, SarJ and SarK, *E. coli* BL21 (DE3)* strain was used for expression. The corresponding plasmid vectors for recombinant protein expression include pET24b and pLATE52, both of which are controlled under inducible T7 promoters. Hence, for heterologous expression of the entire *sar* BGC, we employed Duet vectors (Novagen) featuring T7 promotors to enable the co-expression of multiple enzymes. The commonly used vectors include pRSFDuet-1, pETDuet-1, pCDFDuet-1, pCOLADuet-1, and pACYCDuet-1. Each Duet vector carries two T7 promoter-driven operons, and thus is suitable to co-express at least two proteins. The vectors carry compatible replication origins and antibiotic resistance genes to effectively maintain multiple plasmids in a single *E. coli* cell. Notably, the target protein expression level can be affected by plasmid copy numbers, and therefore, it is necessary to optimize the combination of Duet vectors when multiple plasmids are required for *in vivo* pathway reconstitution. A high plasmid copy number may result in extra metabolic burden and reduce bacterial growth rate,[113] and therefore, pRSFDuet-1 should be chosen with caution in a co-expression system. Considering the plasmid copy number and resistance marker, we determined to choose pETDuet-1 and pCDFDuet-1 in a two-plasmid system for heterologous expression.

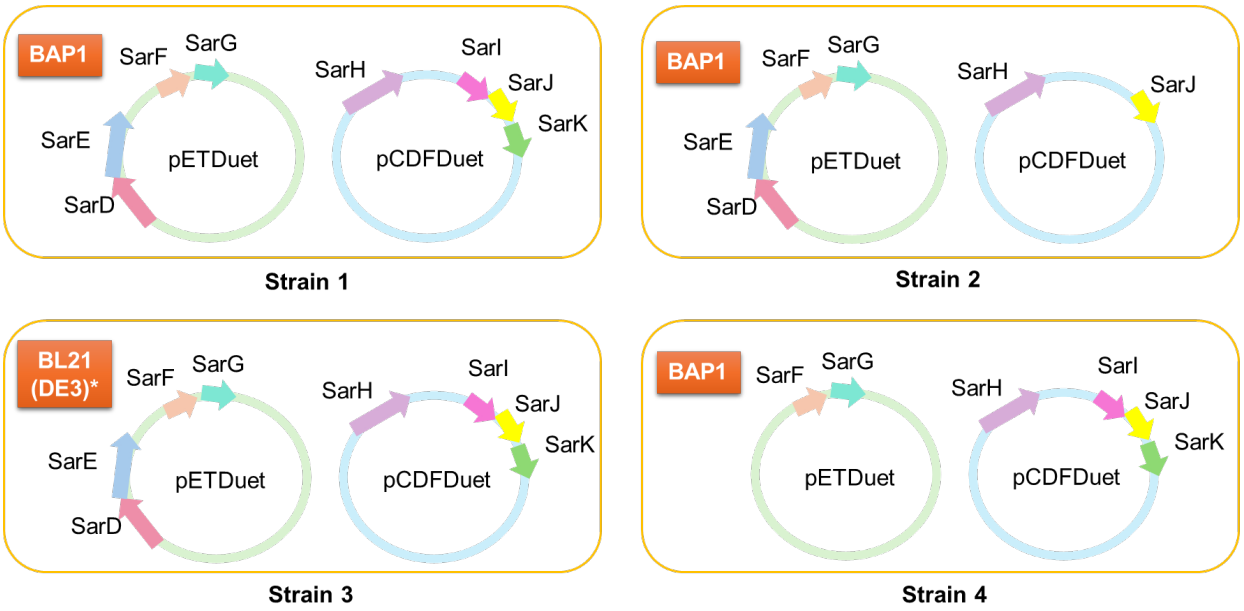


Figure 3-2. Schematic of four strains for heterologous expression of *sar* BGC.

For the initial trial, given the minimal set of biosynthetic enzymes required for *in vitro* production of salivabactin, we inserted genes encoding SarD, SarE, SarF, SarG, SarH and SarJ into pETDuet-1 and pCDFDuet-1 vectors and used *E. coli* BAP1 strain for expression. Then constructs comprising all genes from *sarD* to *sarK*, along with negative controls were prepared for testing with *E. coli* BAP1 and BL21 (DE3)* strains, aiming to identify the optimal combination. In total, we constructed four strains (**Figure 3-2**). Strain **1** and strain **2** were designed to elucidate the impact of SarI and SarK on salivabactin production *in vivo*. Strain **1** and strain **3** were created to assess the potential contribution of the internal *sfp* gene to salivabactin production. Strain **4** served as a negative control. Subsequently, we examined these four strains by growing them first in LB and then transferred to F1 minimal medium for another 48 hours for compound production. F1 minimal medium is used for high-level polyketide production.[114] Since salivabactin is a PK/NRP hybrid, we adopted this medium for potential optimal production. With the detection of LC/HRMS, a semi-quantitative comparison was conducted, revealing that strain **1** exhibited a superior capacity for compound production (**Figure 3-3**). For one thing, the results suggested that the co-expression of *sarDEFGHJ* was able to produce salivabactin without the need of other enyzmes, reaffirming the minimal set of enzymes. However, with the presence of SarI and SarK, the production of salivabactin was significantly enhanced by 10-fold, in alignment with their functions of type II TE and PPTase. Furthermore, strain **1** produced more salivabactin than strain **3**, indicating that despite the presence of the native PPTase, SarK, the promiscuous PPTase, Sfp, aided in compound production by activating carrier proteins in the assembly line. For another, in F1 minimal medium, natural products typically require a minimum of 48 hours to reach maximum production. Nonetheless, the production of salivabactin remained unchanged between 24 and 48 hours, suggesting F1 minimal medium might not be the optimal medium for salivabactin production. Additionally, through comparative analysis between strain **1** and the native host, we found that the salivabactin titers were comparable, signifying a promising initial stage for the development of a scalable heterologous host system (**Appendix F**). To sum up, based on the comparison of these four strains, we identified strain **1** to be the most favorable candidate for subsequent optimization and engineering to increase compound titer.

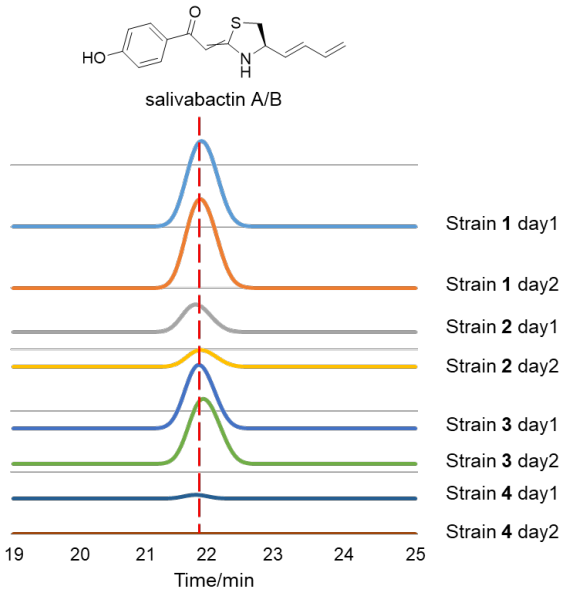


Figure 3-3. Comparison of salivabactin production from four strains, measured by LC/HRMS.

3.2.2 Optimization of growth conditions

From the preliminary experiments of heterologous expression in F1 minimal medium, we established the platform for further titer enhancement. Medium optimization is the first problem to be addressed. We realized that salivabactin production lacks repeatability in F1 minimal medium which, moreover, presents practical challenges due to its labor-intensive nature. Therefore, LB, tryptic soy broth (TSB) and terrific broth (TB) media were tested with strain **1** for salivabactin production. Besides media, it is also crucial to evaluate the impact of substrates and time course of salivabactin production. As per the elucidated biosynthesis, it has been established that 4HBA and Cys serve as two indispensable substrates for the assembly line *in vitro*. Thus, after inoculation, strain **1** grew in LB/TSB/TB medium. Upon OD_{600} reaching 0.6, induction of biosynthetic enzyme expression was initiated by the addition of 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). Meanwhile, 4HBA and Cys were supplemented to the cell culture to facilitate salivabactin production. To evaluate the impact of feeding substrates, we conducted comparative analysis by omitting either 4HBA, Cys, or both substrates from the culture. During growth, samples were extracted by ethyl acetate at intervals of 6 hours, 18 hours, and 25 hours to meticulously examine the time course of production. Based on the detection of LC/HRMS EIC, TB was the optimal medium for salivabactin production and samples extracted after 18h of growth contained the most salivabactin (**Figure 3-6**, **Appendix G**). Although TB is a nutrient-rich medium, it remains less complex compared to BHI medium. Additionally, *E. coli* exhibits a cleaner metabolite background than SAL. Thus, salivabactin purification from extracts derived from strain **1** in TB medium will be more straightforward. In LB samples, the presence of 4HBA was found to be important for achieving significant levels of salivabactin production, while Cys did not exhibit a comparable level of importance in this regard. Compared with the native host, SAL, the titer of salivabactin in *E. coli* cultures was increased by 25% on the basis of EIC MS peak area (**Appendix G**). We therefore focused on optimizing the growth conditions of strain **1** in TB.

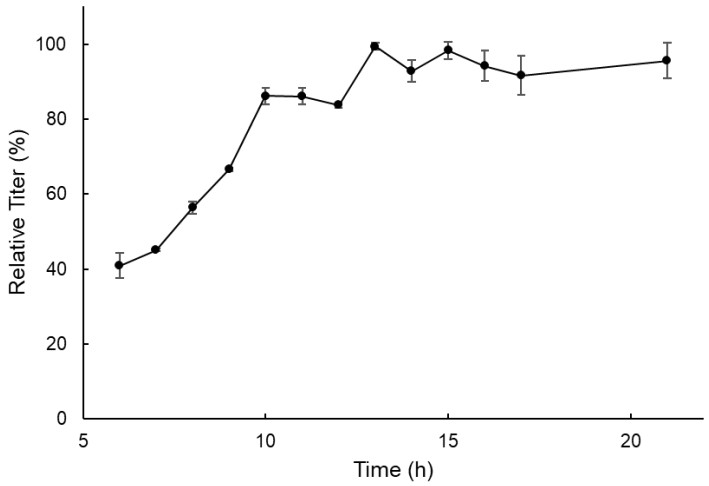


Figure 3-4. Time course for salivabactin production from strain **1** in TB medium with 0.5 mM IPTG and 0.5 g/L 4HBA at 18 ºC, measured by UV-vis spectroscopy. Data are shown as mean ± s.d. (n = 3).

Considering that MS EIC may not offer complete quantitative accuracy and acknowledging the strong absorbance of salivabactin at 350 nm, we transitioned to the quantification of ultraviolet–visible (UV-vis) spectroscopy. One primary challenge associated with SAL is the transient nature of salivabactin production in cultures. Thus, the foremost consideration is to investigate the stability of compound production by strain **1**. With the addition of 0.5 mM of IPTG

and 0.5 g/L of 4HBA in TB medium, the time course of salivabactin production was monitored by UV-vis spectroscopy (**Figure 3-4**). It is apparent that salivabactin maintains stability in *E. coli* cultures with no significant alterations in titer observed between the time points of 13 hours and 21 hours. *E. coli* BAP1 demonstrated notable proficiency as a heterologous host for the expression of the *sar* BGC.

In addition, we scrutinized the impact of fed substrates on salivabactin production (**Figure 3-5**). The pinnacle of salivabactin production in the SAL cultures is attained through the supplementation of 0.5 g/L of Cys. Thus, we initiated our experiments in *E. coli* cultures by adding 0.5 g/L of Cys as well. It turned out that the supplementation of Cys decreased the production of salivabactin in *E. coli* whereas the addition of 4HBA remarkably augmented production levels. The disparity in the effect of Cys on salivabactin production between SAL and *E. coli* may be attributed to variations in medium composition and host metabolism. Therefore, for the following tests, we only supplemented 4HBA in *E. coli* cultures.

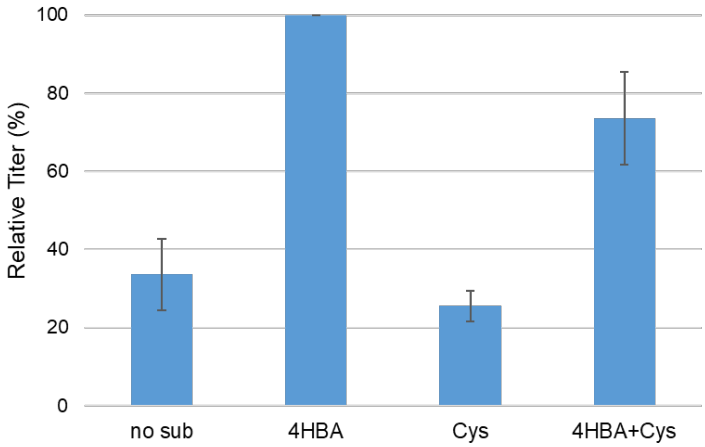


Figure 3-5. Impact of substrate feeding on salivabactin production from strain **1** in TB medium with 0.5 mM IPTG, measured by UV-vis spectroscopy. The tested concentrations of 4HBA and Cys were both 0.5 g/L. Data are shown as mean ± s.d. (n = 3).

Following the confirmation of salivabactin stability and the determination of optimal medium and substrate feeding conditions, our investigation extended to explore additional factors such as induction temperature, growth duration, and concentrations of IPTG and 4HBA. More importantly, we aimed to conduct a comparative analysis between the optimized salivabactin production in *E. coli* and the native host, SAL, with the objective of assessing the magnitude of improvement achieved. After testing various growth conditions (**Figure 3-6**), we found that 0.1 g/L of 4HBA and 0.1 mM of IPTG supplemented when induction could yield higher salivabactin production compared to 0.5 g/L of 4HBA and 0.5 mM of IPTG. Using a lower post-induction temperature, such as 18 ºC, yielded superior results compared to higher temperatures. Nonetheless, reducing the IPTG concentration slowed the protein expression rate, necessitating an extended duration for salivabactin production. Consequently, a duration of 16 hours exhibited an increase in production titer by approximately 40% compared to a 13-hour duration. In summary, our investigation identified optimized growth conditions, wherein strain **1** grew in TB medium with the induction of 0.1 mM of IPTG and the feeding of 0.1 g/L 4HBA. Following induction, strain **1** went through an additional 16-hour growth period at 18 ºC. Remarkably, under these optimal conditions, the salivabactin titer experienced a seven-fold enhancement compared to the native

host SAL cultivated in BHI medium. According to the standard curve crested by UV-vis spectroscopy (**Appendix H**), the compound titer in heterologous expression reached 1.0 mg/L.

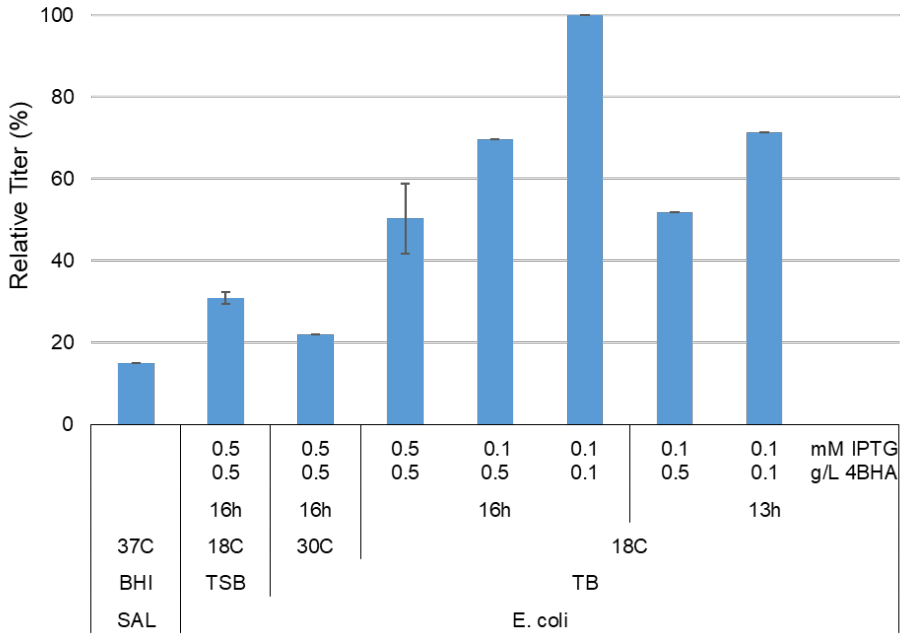


Figure 3-6. Titer comparison of salivabactin from different hosts and different growth conditions, measured by UV-vis spectroscopy. Relative titer was calculated by integrated UV 350 nm peak area.

3.2.3 Starter unit engineering

Considering the potent antimicrobial activities of salivabactin, we aimed to generate various salivabactin analogs with the *E. coli* heterologous expression system and evaluate their bioactivities to facilitate further studies on their mode of action. Given that aromatic acid-activating A domains are typically versatile in substrate recognition, we employed precursor-directed biosynthesis (PDB) for starter unit engineering to generate potential analogs. Various starting materials, including 4-hydroxybenzoic acid (4HBA), 4-aminobenzoic acid (4ABA), benzoic acid (BA), 4-chlorobenzoic acid, 4-azidobenzoic acid, and gallic acid (3,4,5-trihydroxybenzoic acid), were supplied in mixture to cultures of strain **1** and strain **5**. The only difference between strain **1** and **5** was the absence of SarI in strain **5**, as SarI, a type II TE, may play a role in substrate correction or proofreading. Hence, both strains were tested for the production of potential analogs. The results from strains **1** and **5** were highly similar. Apart from the native substrate, 4HBA, we observed the incorporation of 4ABA and BA to yield compound **4** and **5** **(Figure 3-7)**, respectively, supported by molecular networking analysis and MS/MS fragmentation patterns (**Appendix I**). The production titer of compound **4** was found to be comparable to that of salivabactin, exceeding ten-fold the quantity observed for compound **5**. Notably, the *in vitro* system we established in biosynthesis reconstitution can generate corresponding standards for cross-validation with *in vivo* results. We thus tested 4ABA and BA as substrates in the enzymatic reactions involving SarDEFGHJ, observing the production of compound **4** and **5**, in alignment with the *in vivo* feeding results. The substantial production of compound **4** from the *in vivo* heterologous expression system suggested the potential for purifying a sufficient quantity of this compound for structural confirmation with nuclear magnetic resonance

(NMR) spectroscopy and antimicrobial assay testing. Additionally, a methylated derivative of salivabactin was detected (**Appendix I**), probably attributed to the promiscuous methyl transferase activity in *E. coli*. Other precursors such as 4-chlorobenzoic acid, 4-azidobenzoic acid, and gallic acid were not incorporated into the assembly line, possibly due to their large size hindering binding to the A domain of SarD. The starter unit engineering approach was based on the inherent promiscuity of the A domain of SarD, suggesting that further pathway engineering could be achieved by engineering the A domain to broaden its substrate scope. Subsequent purification of these analogs will be conducted for bioactivity assessment. While 4HBA serves as a natural metabolite of *E. coli* and a crucial precursor of Coenzyme Q10 (CoQ10) for bacterial survival, efforts are underway to overexpress the exporter of 4HBA to mitigate competition between the native substrate and fed precursors.

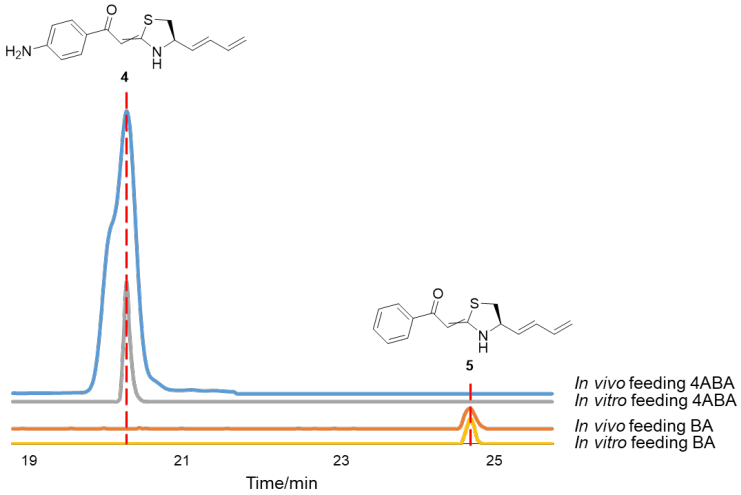


Figure 3-7. Results of *in vivo* and *in vitro* starter unit engineering, measured by LC/HRMS-EIC.

3.3 Discussion

The substantial increase in salivabactin production in *E. coli* demonstrates the robustness of heterologous expression. Heterologous expression technologies involving diverse strains across bacteria, yeasts and filamentous fungi are widely used in both laboratory research and industrial applications.[115,116] The selection of heterologous hosts is paramount for achieving successful heterologous expression outcomes. When deliberating on a heterologous expression system, the following general criteria should be considered: (i) requirement of translational modification, cofactors and processing of biosynthetic enzymes; (ii) authenticity of expressed proteins; (iii) protein expression level and scalability; (iv) cell-associated or secreted products; (v) cytotoxicity of products; and (vi) cost, safety, and other regulatory-related issues.

E. coli stands out as our choice for several reasons. Its suitability arises from its facile genetic manipulation, rapid growth rate, minimalistic media requirements, and cost-effectiveness. Additionally, the successful expression of all required enzymes in *E. coli* during the process of biosynthesis elucidation further ensures the feasibility of utilizing heterologous expression to produce salivabactin. However, despite the advantages above, *E. coli* presents certain drawbacks, including the lack of posttranslational modifications, the propensity for protein aggregation leading to the formation of inclusion bodies, and the inability to secrete certain products into the media. To address these limitations, firstly, we employed the BAP1 strain to facilitate the posttranslational modification of carrier proteins, ensuring their proper functionality. Secondly,

while high-level expression of recombinant proteins in *E. coli* is often desirable for enhanced compound production, it can also result in detrimental effects such as cellular damage, increased metabolic burden, and the production of misfolded proteins. Striking a balance between achieving sufficient enzyme expression for biosynthesis and avoiding cellular stress is crucial. To this end, we meticulously selected and compared the vectors utilized in the *sar* BGC heterologous expression system. Furthermore, continuing endeavors in metabolic engineering are currently directed towards fine-tuning the impact of various promoters. Lastly, salivabactin exhibits potent antimicrobial activity against gram-positive bacteria. *E. coli*, a gram-negative bacterium, was thus thought to be an ideal host for compound production. However, the consistent observation of lysed cells during cell harvesting indicates a potential intracellular cytotoxicity of salivabactin. This underscores the challenge of developing antibiotics for gram-negative bacteria due to their protective outer membrane. Additionally, this observation stimulates our interest in the mode of action of salivabactin. This observation also provides insights into engineering strains to facilitate the export of salivabactin, potentially mitigating cytotoxicity. This strategy could prolong cell viability and consequently enhance compound production.

Each expression system offers a unique set of properties, and successful production of a heterologously produced compound often entails customizing the system based on the properties of the biosynthetic enzymes and the final compound. This necessitates extensive optimization of various components, including medium composition, induction, temperature, duration, and the fed substrates. The heterologous expression system we employed features a *lac*-IPTG induction mechanism. By adjusting the concentration of IPTG, expression of proteins can be regulated at different levels. Lower-level expression can increase the solubility and activity of some target proteins.[117] Hence, we observed higher production of salivabactin with less IPTG supplemented. Nonetheless, the optimal growth conditions vary based on enzyme and product properties. To leverage heterologous expression systems, fine-tuning of these growth parameters is crucial.

In conclusion, after comprehensively delineating the biosynthesis of salivabactin, we embarked on the engineering of its biosynthetic pathway to enhance production. Considering the limited and transient production of salivabactin in the native host, we turned to heterologous expression to boost production. The *E. coli* BAP1 strain was selected as the heterologous host, where we observed stable salivabactin production. Through extensive optimization of growth conditions, we achieved a seven-fold increase in compound titer compared with that in the native host. Furthermore, considering the potent bioactivity of salivabactin, we generated analogs with the aim to study their antimicrobial activities, which facilitated the understanding of mode of action of salivabactin. These analogs were crafted via precursor-directed biosynthesis and starter unit engineering strategies. Overall, our ability to transfer a complex biosynthetic pathway to a more genetically tractable and readily scalable strain represented a significant advancement in the heterologous production of PKs/NRPs endowed with potent bioactivities.

3.4 Materials and Methods

Chemicals and reagents. The chemical, 4-hydroxybenzoic acid, was purchased from TCI Chemicals. Todd-Hewitt broth (THB) and Brain Heart Infusion Broth (BHI) were purchased from BD Biosciences. Other chemicals were purchased from Fisher Scientific or Sigma-Aldrich. Phusion High-Fidelity PCR Master Mix (NEB) was used for PCR reactions. Restriction enzymes were purchased from Thermo Scientific. NEBuilder HiFi DNA Assembly Master Mix was used for DNA construct cloning.

Table 3-1. Oligonucleotides used in this study.

Primers	Sequence (5' → 3')	Notes
PDG62_F	CTTTAAGAAGGAGATATACCATGGGAGAATTTTATTCAAAAATAAG	Co-expression of SarDE and SarFG
PDG66_R	AGCATTATGCGGCCGCAAGCTTATTCATTTACTATTCCTCCTAATAC	
PDG67_F	TATAAGAAGGAGATATACATATGAATGATAATACAGGTTACATC	
PDG67_R	GTTTCTTTACCAGACTCGAGTTATAATTCATCTAAAACAACCTTTC	
PDG68_F	TAAGGAGATATACCATGGGCAATTATAATGATGTTAAAAAAATTATTTA	Co-expression of SarH and SarJ
PDG68_R	ATTATGCGGCCGCAAGCTTAATCCTCCTTAAAAAAATTACTTATT	
PDG63_F	TATAAGAAGGAGATATACATATGGAAAAAACAGTTTTTCTTTTTAC	
PDG63_R	GTTTCTTTACCAGACTCGAGTTAGTTGTTACCTTCAGCTCCAA	
PDG70_F	TATAAGAAGGAGATATACATATGCAGAGTGAGAATAATATTA	Co-expression of SarH and SarIJK
PDG70_R	GTTTCTTTACCAGACTCGAGTCACAATAATTCAGATAAGC	

Table 3-2. Plasmids used in this study.

Plasmid	Vector	Description
pDG36	pETDuet	Co-expression of SarDE and SarFG
pDG39	pCDFDuet	Co-expression of SarH and SarJ
pDG42	pETDuet	Expression of SarFG
pDG43	pCDFDuet	Co-expression of SarH and SarIJK
pDG57	pCDFDuet	Co-expression of SarH and SarJK
pYK01	pACYCDuet	Co-expression of SarDE and SarFG
pYK02	pRSFDuet	Co-expression of SarDE and SarFG
pYK03	pCDFDuet	Co-expression of SarDE and SarFG
pYK04	pETDuet	Co-expression of SarH and SarIJK
pYK05	pACYCDuet	Co-expression of SarH and SarIJK
pYK06	pRSFDuet	Co-expression of SarH and SarIJK

Table 3-3. Bacterial strains used in this study.

Strain	Description
S. salivarius K12	Native host for salivabactin production
E. coli XL1 Blue	DNA cloning and amplification
E. coli BL21 (DE3)*	Recombinant protein overexpression and compound heterologous expression

E.coli BAP1	Recombinant protein overexpression and compound heterologous expression

Construction of plasmids for protein expression in *E. coli*. Individual genes were PCR amplified from the megaplasmid extracted from *S. salivarius* K12. All primers are listed in **Table 3-1**. Gibson assembly was used to ligate PCR products to linear pET24b vector. Thermo Scientific aLICator Ligation Independent Cloning and Expression System was used for plasmid ligation. Plasmids were isolated using a Zyppy Miniprep Kit (Zymo Research) and confirmed by DNA sequencing (UC Berkeley DNA Sequencing Facility).

Heterologous expression for the production of salivabactin. For the small-scale testing, overnight seed cultures ~3 ml in LB were grown from colonies or glycerol stocks of *E. coli* BAP1 strain, containing pDG36 and pDG39 or pDG43. Inoculate 2 mL seed to 200 mL LB with 100 μg/ml carbenicillin and 50 μg/ml spectinomycin. Monitor OD_{600} and when it reached ~0.6, chill on ice for 10 min. Centrifuge in 50mL falcon tubes x 4 at 3500 g for 10 min at 4 °C. Empty media and cells were resuspend in 40 mL of F1 medium with same antibiotics. IPTG (0.5 mM), 4HBA (1 mM) and Cys-HCl (0.5 g/L) were added to 125 mL flasks with spring in bottom. Grow and shake at 16°C for another 48 h.
The optimized conditions: the E. coli BAP1 strain, containing pDG36 and pDG43 (or other combinations), were grown in 10 mL LB medium for seeding overnight at 37 °C with 100 μg/ml carbenicillin and 50 μg/ml spectinomycin. The seed cultures were inoculated into 1 L TB medium to grow until OD_{600} reached 0.6. Cultures were then cooled on ice for 10 minutes before the addition with 0.1 mM IPTG and 0.1 g/L 4BHA. After 18 h of growth at 18 °C, the culture broth and the cell pellets were separated by centrifugation (6000 x g, 15 min, 4 °C).
Compounds were extracted from cell-free supernatant (1 L) and cell pellets using ethyl acetate. The solvent was removed by rotary evaporation. The residue was redissolved in 1 mL methanol and analyzed by LC/HRMS (Inject 100 μL cell culture equivalent). LC/HRMS analysis was performed on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS. LC-UVMS analysis was performed on an Agilent 6120 Single Quadrupole LC/MS with a 1260 series DAD. Each instrument was equipped with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2–98% CH_3CN (v/v) with 0.1% (v/v) formic acid over 30 min in H_2O with 0.1% (v/v) formic acid at a flow rate of 0.5 ml/min was used. LC-HRMS/MS analysis was performed using the same program described above with untargeted mass/mass and a collision energy of 15 eV.

THY Medium recipe. Todd-Hewitt broth from BD Biosciences was supplemented with 3 g/L yeast extract.

F1 medium recipe. 1L of F1 medium contains 3 g KH_2PO_4, 6.62 g K_2HPO_4, 4 g $(NH_4)_2SO_4$, 150.5 mg $MgSO_4$, 5–20 g glucose, 1.25 mL trace metal solution (optional), 100 μM $Fe(NH_4)_2(SO_4)_2$, and 10 mL 100x vitamin solution (optional). Trace metal solution is composed of 40 mg $ZnCl_2$, 200 mg $FeCl_3 \cdot 6H_2O$, 10 mg $CuCl_2 \cdot 2H_2O$, 10 mg $MnCl_2 \cdot 4H_2O$, 10 mg $Na_2B_4O_7 \cdot 10H_2O$, 10 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ in 1 L of DI water. 100x vitamin solution contains 100 mg/L biotin, 100 mg/L choline chloride, 100 mg/L D-calcium pantothenate, 100 mg/L folic acid, 100 mg/L nicotinamide, 100 mg/L pyridoxal hydrochloride, 10 mg/L riboflavin, 100 mg/L thiamine hydrochloride, 200 mg/L i-inositol, and 8.5 g/L NaCl.[68]

Enzymatic reactions to produce compound 4 and 5. A reaction mixture in 50 μL contained 50 mM HEPES (pH 8.0), 2.5 mM $MgCl_2$, 5 mM ATP, 1mM TCEP, 2 mM 4ABA or BA, 2 mM Cys-HCl, 2 mM NADPH, 1 mM Malonyl CoA, 20 μM SarD, 20 μM SarE, 20 μM SarF, 20 μM SarG and 20 μM SarH. The reaction was initiated by the addition of 10 μM SarJ. The reaction was quenched with the addition of 100 μL cold MeCN. Precipitated proteins were removed by centrifugation. The supernatant was subjected to LC/HRMS analysis with 30 μL injection.

Chapter 4. Biochemical Studies of Terminal Alkyne Formation Enzymes

4.1 Introduction

The discovery of JamB marks a pivotal milestone, propelling extensive exploration in the field of biosynthetic terminal alkyne formation. JamB is a membrane-bound bifunctional fatty acid desaturase/acetylenase found in the biosynthetic pathway of cyanobacterial natural product jamaicamide B, accompanied by a fatty acyl-ACP ligase named JamA and an ACP named JamC. Moreover, although JamA is able to activate fatty acids of different lengths and unsaturation patterns, JamB was demonstrated to have more stringent substrate specificity toward both the acyl chain length and the ACP.[73] JamB has substrate specificity towards C_6-chain length, requires a terminal alkene for acetylenase activity, and exhibits stronger protein-protein interactions with JamC than JamA.[118] These findings underscored the intricate functional dynamics of JamB, illuminating its essential role in the molecular mechanism of terminal alkyne biosynthesis. In addition, the activity of JamB was improved ~20-fold in E. coli by random mutagenesis and an efficient fluorogenic screening platform through click reactions. The optimal JamB mutant is JamB-M5T.[119] Similarly, the homologous desaturase/acetylenase TtuB, revealed from genome mining, showed stringent substrate specificity toward C_{10} fatty acyl moieties.[77] To employ JamB/TtuB to label natural products *in situ* with terminal alkynes, a promiscuous type III PKS from the plant *Huperzia serrate* named HsPKS1 was applied to assemble polyketide backbones using malonyl-CoA extender units.[73] HsPKS1 is able to accept a variety of starter units, including aromatic and aliphatic (C_6–C_{10}) CoA thioesters,[120] making the recognition of the 5-hexynoyl and 9-hexynoyl starter units viable. Therefore, heterologous expression of JamABC/TtuABC and HsPKS1 in *E. coli* yield an *in vivo* activity reporting system to evaluate the performance of JamB/TtuB in terminal alkyne formation. We found that TtuB is a much more efficient bifunctional desaturase/acetylenase than JamB in *E. coli* because TtuB is capable to catalyze the conversion of decanoyl-TtuC to 9-decynoyl-TtuC while JamB can only produce a trace amount of alkyne-bearing PK from the feeding of hexanoic acid.[77] Furthermore, a recent study investigated several biosynthetic engineering strategies to interface the JamABC cassette with modular type I PKSs with the aim of creating a generalized approach for terminal alkyne incorporation into polyketide scaffolds.[78] The production efficiency can be improved by enhancing the interactions between the carrier protein (JamC) and PKSs using docking domains and site-directed mutagenesis of JamC.

Extensive investigations into JamB/TtuB have been previously outlined. As membrane-bound bifunctional desaturase/acetylenases, JamB and its homolog TtuB belong to the family of desaturases that catalyze O_2-dependent dehydrogenation of C-C bonds through a non-HEME diiron dependent mechanism.[75] Despite these insights, a comprehensive characterization of the mechanism remains elusive, leaving many aspects of terminal alkyne formation unexplored, including active sites, binding pockets, catalytic cycles, and other mechanistic intricacies. To advance our understanding and broaden the scope of *in situ* tagging natural products through *de novo* terminal alkyne biosynthesis, further mechanistic studies are imperative.

However, the notorious challenges in working with membrane-bound proteins have impeded the discovery and mechanistic understanding of this family of enzymes. To date, only one membrane-bound desaturase, a mammalian stearoyl-CoA desaturase (SCD1), with a sequence

identity/similarity of 21%/38% to JamB, has been documented to possess crystal structures,[121–123] underscoring the scarcity of structural insights into membrane-bound desaturases and the significance of further structural elucidation efforts in this enzyme family. SCD1 is an integral membrane protein located in the endoplasmic reticulum and catalyzes the formation of a *cis*-double bond between the ninth and tenth carbons of stearoyl- or palmitoyl-CoA. The reaction requires molecular oxygen, which is activated by a di-iron center, and cytochrome b_5, which regenerates the di-iron center.[121] Crystallographic analyses have successfully resolved the structures of mouse SCD1 and human SCD1 complexed with stearoyl-CoA substrate; nonetheless, these structures revealed the presence of two zinc ions, rendering the enzymes biochemically inactive.[121,122] The mouse SCD1 was crystallized in lipidic cubic phase while the human SCD1 was crystallized with detergent. Subsequently, the Zhou group incorporated iron into heterologously expressed proteins in HEK293 cells and solved the crystal structure of mouse SCD1 with its diiron center to 3.5 Å. They established the *in vitro* enzymatic reaction by assembling the entire electron transport chain using the purified soluble domains of cytochrome b5 (cyt b_5) and cytochrome b5 reductase (b_5R), and the purified mouse SCD1.[123]

We understand the necessity of obtaining solubilized and active JamB/TtuB for both biochemical and structural interrogation. The solubilization of membrane proteins requires the overlap of two contradictory events: the extraction of membrane proteins from their native lipid membranes and their subsequent stabilization in aqueous environments.[124] Conventionally, solubilizing membrane proteins prior to biochemical analysis involves the use of detergents at concentrations surpassing their critical micelle concentration. There are no golden rules for selecting suitable detergents for the solubilization of a specific membrane protein. Hence, an exhaustive approach must be employed to identify the optimal detergent for each individual protein, balancing both solubilization efficacy and protein stability. Nevertheless, this is rarely achieved owing to the substantial cost and time demands, therefore, a traditional detergent satisfying most of the required needs is often implemented instead. Additionally, nanodiscs have proven to be an invaluable tool for revealing the structure and function of isolated membrane proteins as well as their complexes with other proteins and lipids.[125] Nanodiscs are discoidal lipid bilayers of 8–16 nm in diameter, which are stabilized and rendered soluble in aqueous solutions by two encircling amphipathic helical protein belts, termed membrane scaffold proteins. The size of nanodiscs is determined by the length of the membrane scaffold protein and the stoichiometry of the lipids used in the self-assembly process. The first nanodisc systems utilized amphipathic, helical belt proteins derived from human ApoAI, the major proteinaceous component of high-density lipoprotein particles, to encapsulate the bilayer section.[126,127] Besides these established methodologies, numerous novel strategies have emerged for solubilizing membrane proteins, including QTY code,[128] random heteropolymers,[129] cell-free protein synthesis[130] and SIMPLEx.[131]

In this chapter, we present our findings on the solubilization of JamB/TtuB with diverse methods, along with assessments of their *in vivo* and *in vitro* activities. Furthermore, we generate homologous models of JamB derived from the crystal structure of SCD1 and propose residues relevant to its acetylenase activity to provide mechanistic insights in terminal alkyne formation.

4.2 Results

In the beginning, we sought to employ detergents for the solubilization of JamB and TtuB; however, our efforts did not yield satisfactory outcomes. We tested a few detergents including n-dodecyl-β-D-maltoside and n-decyl-β-D-maltopyranoside, which showed moderate solubilization

efficiency (**Appendix J**). Nonetheless, the conventional Ni-NTA affinity purification method for His-tagged proteins proved ineffective for detergent-solubilized JamB and TtuB. Alternative affinity chromatography techniques warrant exploration. Notably, *in vitro* activity assays of solubilized JamB/TtuB revealed no detectable terminal alkyne production, suggesting the need for further method investigation. Concurrently, we explored alternative approaches to achieve direct solubilization of JamB/TtuB without the need for subsequent re-solubilization with detergents.

4.2.1 Solubilization by QTY code and biochemical characterization

The QTY code, that is named for three amino acids, glutamine (Q), threonine (T), and tyrosine (Y), represents a robust and straightforward tool for making membrane proteins become water-soluble without detergents.[128] In this approach, the hydrophilic amino acid residues, Q, T, and Y, are used to systematically replace the hydrophobic amino acid residue, leucine (L), valine (V), isoleucine (I), and phenylalanine (F) in transmembrane regions of membrane proteins. Successful solubilization and functional reconstitution were obtained in chemokine receptors through the QTY code engineering. Thus, the QTY code captured our interest due to its capability to enable systematic design of membrane proteins via simple and specific amino acid substitutions.

Based on the predicted membrane topology of JamB by TOPCONS,[73] the first JamB variant was generated following the principle of substituting L with Q, I and V with T, and F with Y in transmembrane sections. A second JamB variant was generated through residue swapping of QTY code in the transmembrane sections predicted by the sequence alignment with the desaturase SCD1. Similarly, TtuB variants were designed based on predicted membrane topology and homologous modelling, respectively. The genes encoding QTY-engineered JamB and TtuB were synthesized and sequenced for recombinant protein purification in *E. coli*. Notably, considering the increased activity of terminal alkyne formation in *E. coli*, we implemented JamB-M5T in all JamB constructs involved in QTY code engineering. Out of the four constructs, only one, which encoded the first JamB variant, yielded a small amount of soluble JamB (37 kDa) with the estimated yield of 0.2 mg/mL (**Figure 4-1**). Nonetheless, it is noteworthy that the majority of engineered JamB remained in the insoluble fraction.

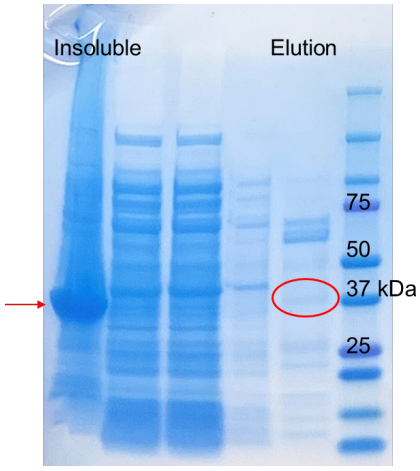


Figure 4-1. Solubilization of JamB by QTY code engineering. SDS-PAGE gel showing a small amount of soluble JamB(QTY). From left to right: insoluble fraction, supernatant, flow-through, wash and elution.

Given the promising solubility outcomes of the engineered JamB construct, we were eager to assess whether it retained its ability to produce terminal alkynes. We started with the *in vivo* activity examination with the established activity reporting system in *E. coli*. Since the substrate

of JamB is tethered with JamC and the formation of terminal alkynes involves a redox reaction requiring electron transport partners, the *in vivo* activity characterization is much easier to conduct than *in vitro* tests. We continued using HsPKS1 as a reporter of engineered JamB activity in *E. coli* systems. Two plasmids encoding *jamAB(QTY)C* or *jamABC* and *hspks1*, with each gene regulated by a T7 promoter, were transformed into the *E. coli* BAP1 strain, yielding the engineered strain **DG2** compared with the stain **DG1** as the positive control (**Figure 4-2. A**). The resulting strains grew in shake flask cultures, and after induction with IPTG, the strains continued to grow in F1 minimal medium supplemented with 5-hexenoic acid at 20 °C for 48 hours. Analyses of these culture extracts were done by LC/HRMS and the standards **6** and **7** were synthesized by enzymatic reactions of 5-hexynoic/hexenoic acid and HsPKS1.[73] The results showed that compound **6** was successfully produced by positive control. However, compound **6** was undetectable in the products of the strain **DG2** from F1 cultures. It is possible that the lack of presence of the alkyne product is due to low titer, rather than complete lack of production. Moreover, the growth medium plays a pivotal role in metabolite production. Therefore, we conducted experiments by consistently growing both strains in LB medium without switching media and subsequently analyzing the culture extracts by LC/HRMS. Surprisingly, we observed the production of both compound **6** and **7** in strain **DG2**, albeit the titer of **6** was approximately one-tenth of that of **7** (**Figure 4-2. B**). We repeated experiments in LB medium and obtained similar results, confirming the *in vivo* activity of QTY engineered JamB.

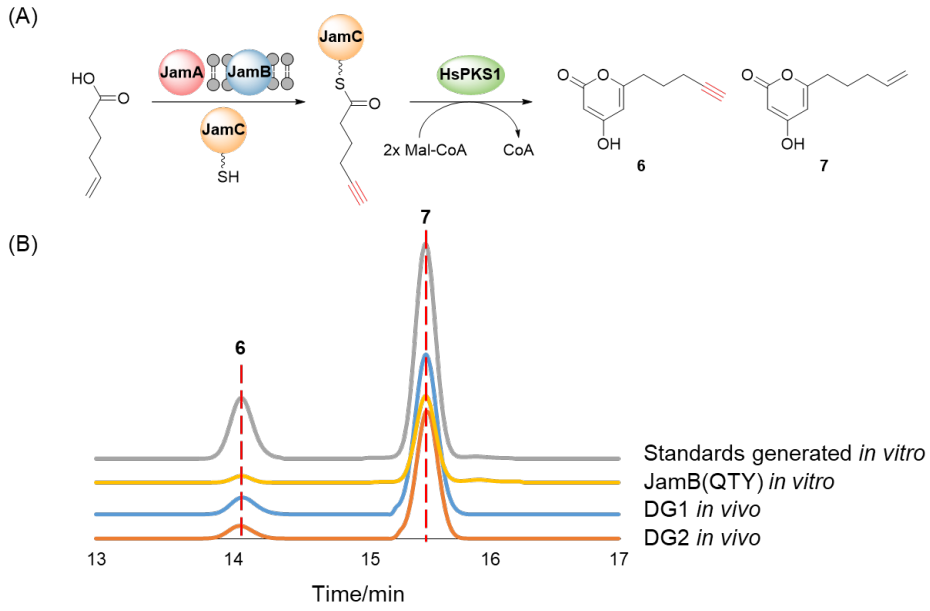


Figure 4-2. Activity tests of JamB(QTY) *in vivo* and *in vitro*. **(A)** The schematic of biosynthesis of alkyne-containing polyketides formation in E. coli. **(B)** Detection of compound **6** and **7** in *E. coli* cultures from LB medium and enzymatic assays involving purified JamB(QTY), measured by LC/HRMS. **DG1** contains WT JamB while **DG2** contains QTY engineered JamB.

Then we proceeded to verify the *in vitro* activity of QTY engineered JamB. Because there was a large quantity of engineered JamB remaining in the insoluble fraction, reliance solely on *in vivo* activity could not ascertain whether the observed activity originated from the soluble or insoluble fraction. To scrutinize the activity of JamB comprehensively, a reliable *in vitro* assay was imperative. The native electron transfer partner is unknown for JamB. We thus added the mixture of NADH, NADPH, ferredoxin (Fx) /ferredoxin reductase (FxR), phenazine methosulfate (PMS)

and ascorbic acid to mimic the native reducing environment. We tested soluble and membrane fractions separated from cell lysate and the purified soluble JamB(QTY) *in vitro* compared with the standards generated by 5-hexynoic/hexenoic acid, JamAC and HsPKS1. We did not observe any terminal alkyne (compound **6**) formation either in membrane or soluble fractions. However, trace amounts of compound **6** were detected in the assay using the purified soluble JamB(QTY), suggesting a potential activity of the soluble JamB(QTY). Further optimizations on increasing the yield of soluble JamB(QTY) and the reaction efficiency of *in vitro* assays are needed. We aim to obtain a sufficient quantity of soluble JamB(QTY) for comprehensive assay studies, and also adjust the assay conditions to make the terminal alkyne formation more detectable, such as using hydroxylamine[132] to hydrolyze molecules off carrier proteins rather than relying on HsPKS1.

4.2.2 Solubilization by CFPS with RHP and biochemical characterization

Given that the QTY code was a radical approach, we pursued other solubilization methods in parallel, as a precaution in case the JamB(QTY) was biologically inactive after altering multiple residues at one time. Random heteropolymers (RHP) reported by the Xu lab is designed to mimic intrinsically disordered proteins for protein solubilization and stabilization in non-native environments. The RHP, with optimized composition and statistical monomer distribution, enables cell-free synthesis of membrane proteins with proper protein folding for transport and enzyme-containing plastics for toxin bioremediation.[129] Cell-free protein synthesis (CFPS) derived from crude cell extracts provides an alternative method for producing proteins that are recalcitrant to recombinant expression *in vivo*, with several distinct advantages. The absence of a cell wall enables an open and versatile environment for active monitoring, rapid sampling, and direct manipulation of the protein synthesis process.[133] Jewett lab reported the production of multicopper oxidases in an *E. coli*-based CFPS system and reconstituted the activity in the following enzymatic assays.[134] In addition to soluble metal-dependent enzymes, they also applied this platform to membrane proteins such as oligosaccharyltransferases.[130] However, there have been no instances of successful production of metal-dependent membrane enzymes in their active forms so far. We thus adopted this strategy and aimed to develop a CFPS system coupled with RHP for the biochemical studies on JamB and TtuB in collaboration with Xu lab.

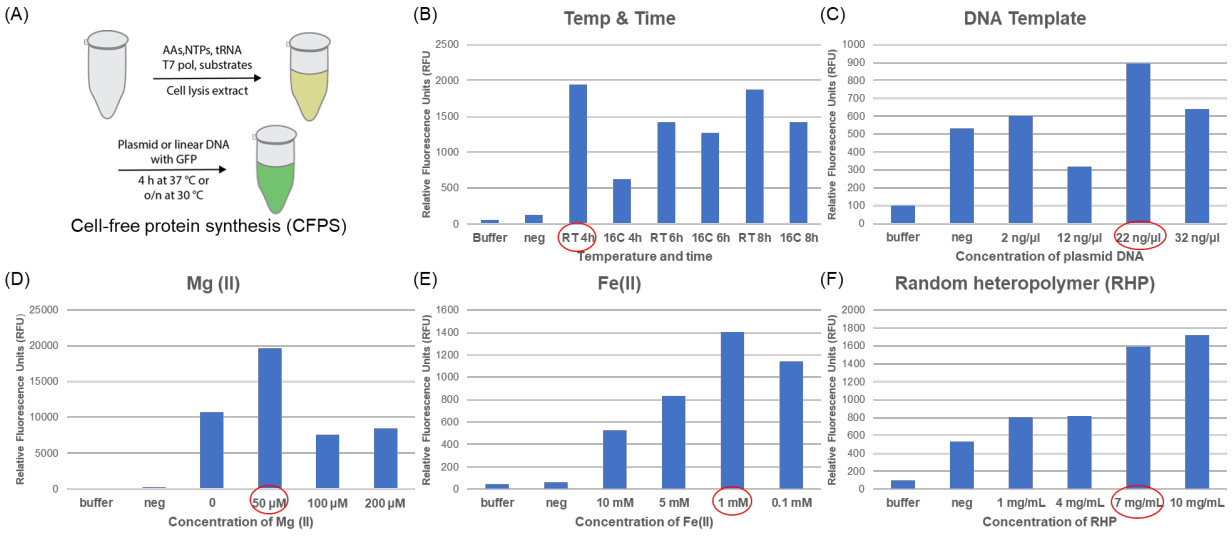
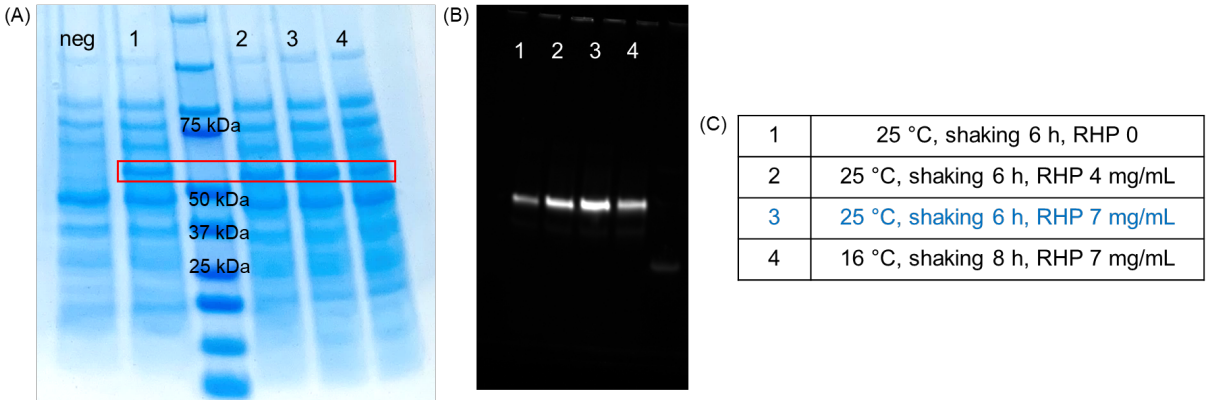


Figure 4-3. First round of CFPS optimization, detected by fluorescence spectroscopy. **(A)** Schematic of CFPS. **(B)** Optimization of incubation time and temperature. **(C)** Optimization of DNA quantity. **(D)** Optimization of Mg (II) concentration. **(E)** Optimization of Fe (II) concentration. **(F)** Optimization of RHP concentration.

We started with optimizations for the expression of JamB in the CFPS system. To monitor the expression level of JamB, we fused superfolder green fluorescent protein (sfGFP)[135] to the C-terminus of JamB for fluorescence detection. This choice was made based on the presumption that the N-terminus of JamB would likely interact with the electron donor.[119] Additionally, sfGFP, as a highly soluble protein, is commonly employed to enhance the solubility of membrane proteins. Moreover, the detection of fluorescence originating from sfGFP suggests potentially proper folding of JamB. Similar to the HsPKS1 activity reporting system in *E. coli*, we used a well-studied type I PKS, LipPKS1,[78] to test the activity of JamB-sfGFP fusion *in vivo* (**Appendix K**). The results indicated that the fused sfGFP did not affect the JamB activity of terminal alkyne formation and the ability to be used by the downstream PKS assembly lines to yield alkyne-containing natural products. Then we moved forward to the utilization of JamB-sfGFP in the CFPS system. To optimize the system, we meticulously adjusted numerous parameters with the quantification of fluorescence intensity, including temperature, incubation time, DNA quantity, and concentrations of Mg (II), Fe (II) and RHP. The optimal conditions identified in this round were incubating 22 ng/μL of plasmid with 50 μM of Mg (II), 1 mM of Fe (II) and 7 mg/mL of RHP at room temperature for 4 hours (**Figure 4-3**).

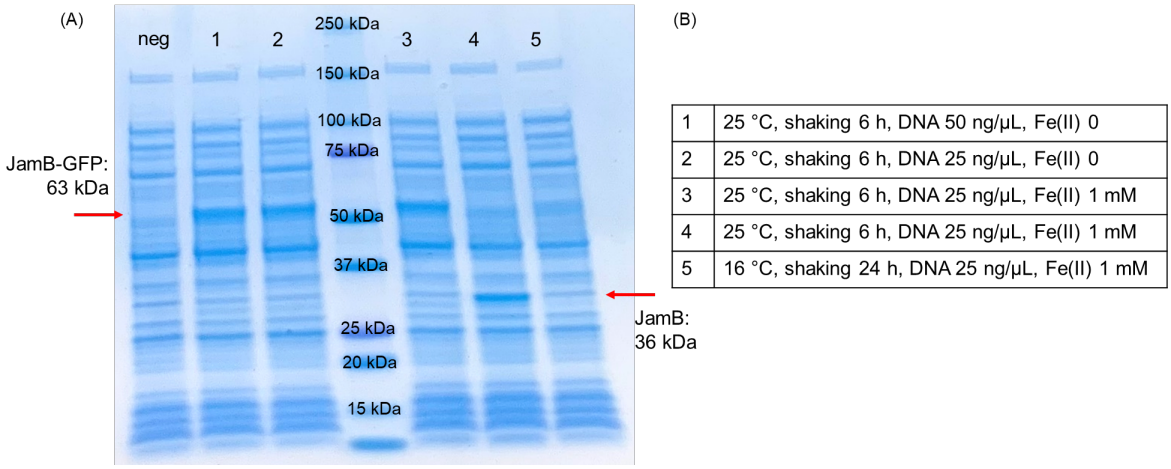


1	25 °C, shaking 6 h, RHP 0
2	25 °C, shaking 6 h, RHP 4 mg/mL
3	25 °C, shaking 6 h, RHP 7 mg/mL
4	16 °C, shaking 8 h, RHP 7 mg/mL

Figure 4-4. Second round of CFPS optimization. **(A)** CFPS of JamB-sfGFP on SDS-PAGE gel under different expression conditions. **(B)** Fluorescence imaging of the SDS-PAGE gel. **(C)** Different conditions tested in the second round of optimization.

With the preliminary optimization, our curiosity led us to investigate whether the complete expression of JamB-sfGFP had been achieved. This prompted us to utilize SDS-PAGE for visualizing protein expression as the second round of optimization. In this round, we employed NEBExpress CFPS commercial kits and implemented shaking of tubes to ensure homogeneous reaction conditions. The negative control excluded the gene encoding JamB-sfGFP, thus presenting only the *E. coli* endogenous proteins. The SDS-PAGE gel clearly demonstrated the successful expression of JamB-sfGFP at the expected size of 63 kDa (**Figure 4-4. A**). Fluorescence imaging of the SDS-PAGE gel provided additional confirmation of the proper folding of sfGFP, thereby suggesting the likelihood of correct folding for JamB as well (**Figure 4-4. B**). More importantly, fluorescence imaging also offered insights into the yield of CFPS under each condition. Even in the absence of RHP, CFPS yielded some expression of JamB-sfGFP, albeit at a low level. However, with the addition of increasing amounts of RHP, fluorescence intensity increased correspondingly,

indicating the assistance of RHP in protein folding, consistent with the reported function. Notably, the protein band looked smaller with the presence of RHP compared to its absence on SDS-PAGE, likely due to its interaction with membrane proteins. Consistent with the first round, CFPS was more efficient at room temperature than at 16 ºC. Hence, the optimal conditions involved shaking tubes containing CFPS reagents at 25 ºC for 6 hours, supplemented with 7 mg/mL of RHP. Then we proceeded with the testing of genes encoding JamB, which resulted in comparable levels of expression of JamB-sfGFP. Furthermore, the addition of Fe (II) did not impact the expression level (**Figure 4-5**). The estimated yield of CFPS for JamB reached 0.5 mg/mL. Overall, we identified the optimal conditions for JamB as shaking tubes containing 25 ng/µL of plasmid with CFPS reagents at 25 ºC for 6 hours, supplemented with 1mM of Fe (II) and 7 mg/mL of RHP.



1	25 °C, shaking 6 h, DNA 50 ng/µL, Fe(II) 0
2	25 °C, shaking 6 h, DNA 25 ng/µL, Fe(II) 0
3	25 °C, shaking 6 h, DNA 25 ng/µL, Fe(II) 1 mM
4	25 °C, shaking 6 h, DNA 25 ng/µL, Fe(II) 1 mM
5	16 °C, shaking 24 h, DNA 25 ng/µL, Fe(II) 1 mM

Figure 4-5. Third round of CFPS optimization. **(A)** CFPS of JamB-sfGFP and JamB on SDS-PAGE gel under different expression conditions. **(B)** Different conditions tested in the third round of optimization.

After the solubilization of JamB with CFPS plus RHP, we applied the optimal conditions to TtuB and observed similar results. For biochemical characterization, we purified JamB and TtuB with Ni-NTA magnetic beads (**Figure 4-6**). Notably, while the expected size of TtuB was 38 kDa, the size observed on the SDS-PAGE gel appeared smaller. We then conducted enzymatic assays with purified JamB-sfGFP and TtuB. For the assay involving JamB-sfGFP, JamAC and 5-hexenoic acid were added. For the assay involving TtuB, TtuAC and 9-decenoic acid were added. Reducing agents such as NADPH and dithionite were supplemented for artificial electron donors. However, terminal alkyne formation was not detected on carrier proteins. Several factors may account for this observation. Firstly, it is possible that the artificial electron donors were not effectively recognized by JamB/TtuB. We plan to explore other reducing agents, such as Fx/FxR and PMS, to address this issue. Secondly, the detection sensitivity of molecules tethered on carrier proteins may be insufficient for detecting low activity enzymes. We intend to investigate alternative methods, such as using chemical reagents or downstream PKS, to offload the small molecules for improved sensitivity. Lastly, it is conceivable that JamB/TtuB solubilized by CFPS with RHP might be inactive. To address this, we will explore CFPS with nanodiscs as an alternative method to preserve active forms.

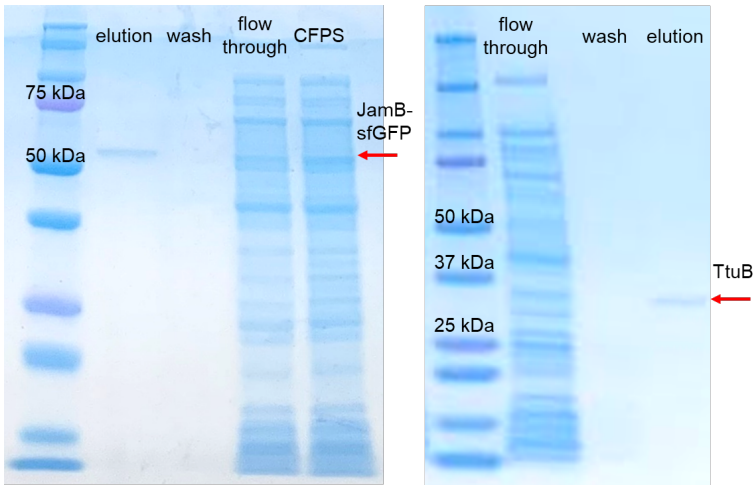


Figure 4-6. Purified JamB-sfGFP and TtuB from CFPS.

4.2.3 Solubilization by SIMPLEx and biochemical characterization

Solubilization of integral membrane proteins (IMPs) with high levels of expression (SIMPLEx) enables soluble expression of IMPs directly in living cells by genetically modifying an IMP target with a truncated version of human apolipoprotein A-I (ApoAI), which effectively shields the IMP from water and promotes solubilization.[131] Truncation variants of human ApoAI lacking its 43-residue globular N-terminal domain (hereafter ApoAI*) have the ability to form nanodiscs into which detergent-solubilized IMPs can partition.[136] In addition to being amphipathic, ApoAI* exhibits substantial structural flexibility that allows it to readily conform to a spectrum of different geometries as needed. More importantly, given the water-soluble IMPs are expressed inside cells, the SIMPLEx method has the potential to enable in vivo applications, for example, metabolic-pathway engineering, that leverage the activity of solubilized IMPs. Considering all the above advantages, we attempted to employ SIMPLEx method for JamB/TtuB solubilization.

The construct used in SIMPLEx method consisted of a highly soluble protein, membrane protein itself, and ApoAI*. The highly soluble protein at the N-terminus was designed to prevent the resulting protein chimera from being inserted into the cytoplasmic membrane. Here, we tested GFP and *E. coli* maltose-binding protein (MBP) as the N-terminal fusion in *E. coli* recombinant protein expression. The constructs comprising MBP-JamB, MBP-TtuB, GFP-JamB and GFP-TtuB did not yield soluble proteins, suggesting that fusion with a highly soluble protein at N-termini alone is insufficient to render IMPs soluble. Similarly, GFP-JamB-ApoAI* and GFP-TtuB-ApoAI* were insoluble. However, we obtained soluble proteins of MBP-JamB-ApoAI* (101 kDa) and MBP-TtuB-ApoAI* (103 kDa) (**Figure 4-7**). Expression of these constructs in TB cultures, compared with LB cultures, slightly increased the yield to 2 mg/mL and 2.4 mg/mL, respectively. The yield of solubilized JamB/TtuB was the highest among the three tested strategies, suggesting that the SIMPLEx method has the potential to produce a sufficient quantity of JamB/TtuB for biochemical scrutiny and structural elucidation. Notably, membrane proteins frequently do not move according to the expected molecular weight in SDS-PAGE. They often move faster possibly due to incomplete unfolding or due to binding more SDS per mass unit protein as compared with a soluble protein. Therefore, both MBP-JamB-ApoAI* and MBP-TtuB-ApoAI* appeared smaller on the SDS-PAGE gel.

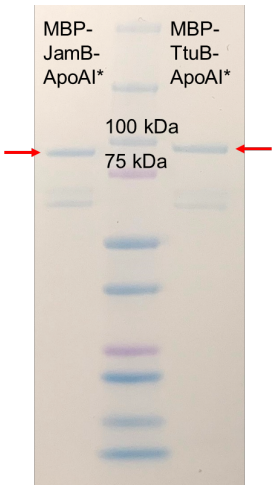
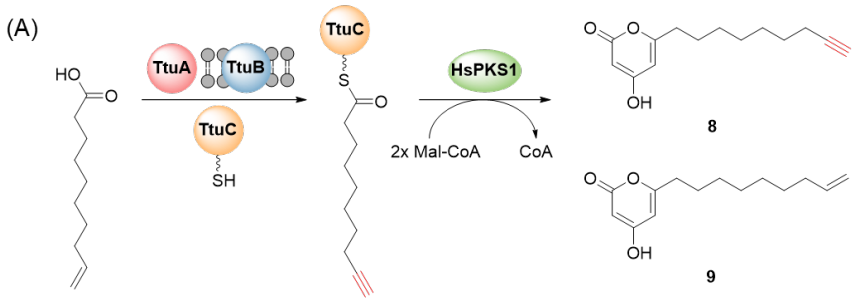


Figure 4-7. SDS-PAGE gel of solubilized JamB/TtuB with SIMPLEx method.

Then we examined the *in vivo* activity of MBP-JamB-ApoAI* and MBP-TtuB-ApoAI* with the *E. coli* HsPKS1 reporting system. The solubilized JamB fusion showed minimal activity, with a 100-fold decrease in terminal alkyne production (**Appendix L**). The N-terminus of JamB was proposed to interact with electron donors for alkyne formation, which might be the reason for the significant decrease in activity. Considering the enhanced activity of JamB(M5T) compared to WT JamB, we intend to evaluate MBP-JamB(M5T)-ApoAI* in future experiments to assess its potential for improved activity. In contrast, the solubilized TtuB fusion exhibited modest activity, with a 10-fold decrease in the production of terminal alkyne containing compound **8** (**Figure 4-8**) compared with WT TtuB. Two plasmids encoding *ttuAB(MBP-ttuB-ApoAI*)C* or *ttuABC* and *hspks1*, with each gene regulated by a T7 promoter, were transformed into the *E. coli* BAP1 strain, yielding the engineered strain **DG4** compared with the stain **DG3** as the positive control. Strain **DG5** contained plasmids encoding *ttuAC* and *hspks1* as the negative control. The *in vivo* activity of MBP-TtuB-ApoAI* was validated with three independent tests. Our next step was to assess the *in vitro* activity of MBP-TtuB-ApoAI*. We conducted enzymatic assays involving purified MBP-TtuB-ApoAI*, TtuA, TtuC, HsPKS1, 9-decenoic acid, cofactors and reducing agents. Unfortunately, no terminal alkyne formation was detected from *in vitro* assays. Notably, solubilized TtuB displayed activity *in vivo* when cultivated in F1 medium supplemented with Fe (II). Therefore, to achieve activity *in vitro*, it may be necessary to express solubilized TtuB under the same growth conditions, especially with extra Fe (II) supplemented, and subsequently purify it. Another possible reason for the lack of activity *in vitro* could be that the shielding ability of ApoAI* also impacted the interaction between TtuB and TtuC.



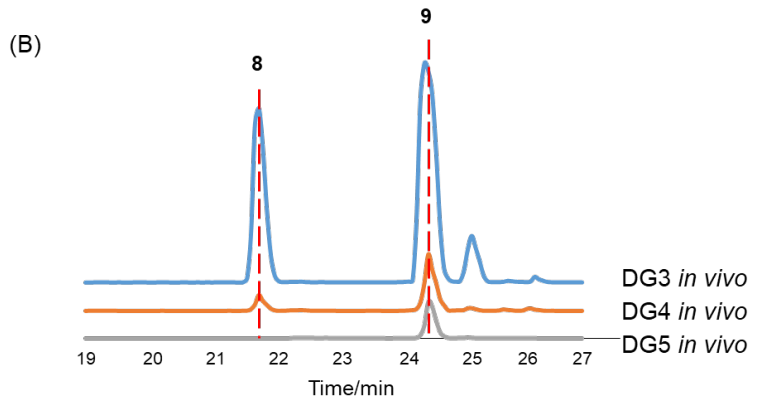


Figure 4-8. Activity tests of MBP-TtuB-ApoAI* *in vivo*. **(A)** The schematic of biosynthesis of alkyne-containing polyketides formation in *E. coli*. **(B)** Detection of compound **10** and **11** in *E. coli* cultures from F1 medium, measured by LC/HRMS. Strain **DG3** was the positive control with WT TtuB. Strain **DG4** contained engineered MBP-TtuB-ApoAI*. Strain **DG5** was the negative control omitting TtuB.

Based on the biochemical studies conducted thus far on solubilized JamB/TtuB, the QTY code method emerged as the most promising approach for JamB, while SIMPLEx method shows the most promise for TtuB, laying the foundation for exhaustive characterization of JamB and TtuB in the future.

4.2.4 Structural modeling of JamB and mechanistic insights

Based on the crystal structure of SCD1, structural modeling of JamB was established by the Robetta web server by William Porterfield, a previous postdoctoral researcher in our lab. The studies on the structure of SCD1 identified the possible residues in the substrate binding channel and the highly conserved nine histidine residues. More importantly, the potential two metal ions are separated by 6.4 Å, which is larger than the distance between common desaturase diiron sites, 4.2 Å. This unusually long distance may play a role in a new mechanism of oxygen activation and metal intermediates in the process of desaturation. According to structural modeling, sequence alignment and structure-function relationships, we identified residues conserved near the active site of JamB, including Q269, N271 and P136 (**Figure 4-9**), which may be important for desaturase versus acetylenase activity. Site-directed mutagenesis and heterologous expression will be performed to test the activity of mutants Q269R, N271S and P136A. In addition, the structure of SCD1 indicates a plausible pathway for lateral substrate access and product egress, and the geometry of the substrate tunnel and the conformation of the bound acyl chain provide a basis for the regioselectivity and stereospecificity of the desaturation reaction. Comparing these features to the modeled JamB structures, we identified two regions at the entrance of the active site of JamB that may be important for ACP recognition (**Figure 4-9**). One region covers two sides of the triangular active site entrance and is approximately 25 amino acids in length and is partially an unstructured loop and partially an alpha helix. The other region is an alpha helix ~10 amino acids long and covers the other side of the entrance. By swapping these regions between JamB and TtuB, we may be able to change the acyl-ACP recognition of each protein. The acyl chain tunnels are well defined with hydrophobic residues lining the walls until the predicted diiron center (**Figure 4-9**). The tunnel is not well conserved between JamB and TtuB, suggesting that swapping some of these residues could change the substrate preference of the enzymes, and engineering of the tunnel can further expand the substrate scope for broader applications.

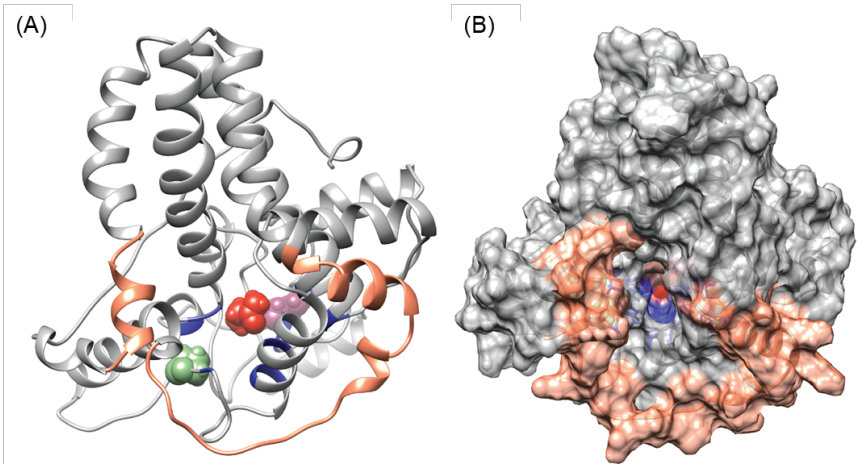


Figure 4-9. Proposed JamB structural model. **(A)** Ribbon model of JamB with important residues shown as spheres (P136 green, Q269 red, and N271 purple). The proposed ACP-binding site is highlighted in orange and histidine residues involved in iron binding are shown in blue. **(B)** Surface of modeled JamB shown from the same view as (A). The substrate tunnel is in the center of the structure with Q269 in red visible near the back. Models were created using the Robetta web server using SCD1 as a template.

4.3 Discussion

Desaturases have independently evolved twice. Integral membrane desaturases are widespread in both prokaryotes and eukaryotes, residing in endomembrane systems, while soluble acyl-ACP desaturases are found in the plastids of higher plants.[75] JamB and TtuB stand out for several reasons. Firstly, their substrates are tethered to an ACP before desaturation, which is unusual for membrane-bound desaturases that typically accept acyl-CoA or glycerolipid substrates. Secondly, instead of generating internal alkenes like most desaturases, JamB and TtuB target the last two carbons in the fatty acyl chain to yield a terminal alkene. Remarkably, it further recognizes a terminal alkene to produce a terminal alkyne. Thirdly, the mechanism by which an acetylenase abstracts two hydrogens from adjacent carbon atoms linked by a double bond without forming an epoxide remains an intriguing mechanistic question. The inherent challenges in studying membrane-bound proteins have hindered the discovery and mechanistic understanding of this enzyme family. Moreover, desaturation occurring on carrier protein-tethered substrates adds another layer of complexity to the process.

SCD1, a membrane-embedded iron dependent enzyme that catalyzes the formation of a double bond on a saturated acyl-CoA, was solubilized with *n*-decyl-β-D-maltopyranoside. The activity was reconstituted with the electron transfer partner, NADH/cyt b5/b5R.[123] Another example is UndB, an iron-dependent integral membrane protein that converts fatty acids to terminal alkenes in the fatty acid desaturase like superfamily. UndB was solubilized with non-ionic detergents, *n*-dodecyl-β-D-maltopyranoside and lauryl maltose neopentyl glycol. Both detergents yielded similar levels of solubilization; however, the latter one solubilized UndB exhibited higher activity. The activity of UndB was reconstituted with NADPH/Fx/FxR. Researchers tested several Fx/FxR derived from different native sources and identified the optimal pair.[137] The preliminary assessment of electron transfer partners of JamB was conducted with semi-purified membrane fractions, revealing that the introduction of NADPH/Fx/FxR increased the activity of JamB by 35%. Although the result is consistent with observations that in cyanobacteria, membrane-bound desaturases typically use Fx instead of cyt b5 as an electron transport partner, and that Fx is often

a required cofactor for soluble acyl-ACP desaturases,[73] further verification will be carried out with purified and solubilized JamB/TtuB.

Given the substrates of JamB/TtuB are conjugated to carrier proteins, there are multiple approaches to detect the conversion from substrates to products. For one thing, the most direct method is the native MS detection of carrier protein bound substrates. Due to the slight mass changes from double bonds to triple bonds compared to overall mass of associated proteins in the case of *jam*/*ttu* systems, the PPant ejection assay allows the mass of substrates loaded onto carrier proteins to be readily deduced from the mass of the corresponding PPant fragments, facilitating sensitive detection of desaturation reactions.[138] For another, substrates tethered to carrier proteins can be offloaded with enzymes or chemicals to yield small molecules for detection. The enzyme we used in *jam*/*ttu* systems is HsPKS1. One advantageous aspect of HsPKS1 is its relaxed substrate specificity, enabling the acceptance of acyl chains from JamC/TtuC without significant preferences. Thus, the production of alkyne or alkene containing polyketides can effectively reflect the activity of JamB/TtuB. However, a drawback is that the turnover rate of HsPKS1 is relatively slow, which requires a long period of incubation time. Alternatively, chemical probes can mitigate the downside. Hydroxylamine is widely used for profiling *S*-acylated fatty acids on proteins.[132] The limitation of using hydroxylamine is that there are no turn overs for carrier proteins, necessitating ample amounts of carrier proteins as substrates in the beginning of assays.

In conclusion, to facilitate *in vitro* biochemical characterization and structural elucidation, we aimed to obtain solubilized and active forms of JamB/TtuB. We employed four strategies including the use of detergents, the QTY code, CFPS with RHP and the SIMPLEx method. We focused on the discussion of the latter three strategies in this dissertation. While all variants of JamB/TtuB exhibited respectable solubility, only JamB derived from the QTY code displayed both *in vivo* and potential *in vitro* activity, and TtuB from the SIMPLEx method exhibited *in vivo* activity. The efforts in solubilization provide a promising foundation for future investigations. Furthermore, the interaction between JamB/TtuB with their respective ACPs plays an important role in substrate recognition. structural modelling of JamB based on the homologous SCD1 structure offered deeper mechanistic insights, thereby facilitating point mutations and residue swapping to further probe the alkyne biosynthesis process.

4.4 Materials and Methods

Chemicals, reagents and strains. Chemicals including 5-hexenoic (purity 99%) and 5-hexynoic (purity 97%) acids were purchased from Alfa Aesar. CoA and Mal-CoA were purchased from CoALA Biosciences. Other chemicals were purchased from Fisher Scientific or Sigma-Aldrich. Phusion High-Fidelity PCR Master Mix (NEB) was used for PCR reactions. Restriction enzymes were purchased from Thermo Scientific. NEBuilder HiFi DNA Assembly Master Mix was used for DNA construct cloning. NEBExpress Cell-free E. coli Protein Synthesis System was used for CFPS. Bacterial strains involved in this study include *E. coli* XL1 Blue for DNA cloning and amplification, *E. coli* BL21 (DE3)* for recombinant protein overexpression and CFPS cell lysate preparation, and *E.coli* BAP1 for recombinant protein overexpression and compound heterologous expression.

Table 4-1. Oligonucleotides used in this study.

Primers	Sequence (5’ → 3’)	Notes

PDG26_R	GACATATGTATATCTCCTTCTTAAAGTTAAACAAA	CFPS of JamB (C-His)
PDG27_F	TTAACTTTAAGAAGGAGATATACATATGTCAATGCCAATGGATGTGAG	
PDG27_R	CGACTTAGTGGTGATGATGGTGATGAGCTAACTTCTTAGCTTCGATCA	
PDG34_F	GAGATATACATATGTCAATGCCAACCG	CFPS of JamB-sfGFP (C-His)
PDG34_R	GACTTAGTGGTGATGATGGTGATGCGTAATACCTGCCGCATTCA	
PDG35_F	ATTACGCATCACCATCATCACCACTAAGTCGACCGGCTGCTAA	
PDG35_R	TGGCATTGACATATGTATATCTCCTTC	
PDG43_F	GTCGACCGGCTGCTAAC	CFPS of TtuB (C-His)
PDG43_R	ATGTATATCTCCTTCTTAAAG	
PDG46_F	GAAGGAGATATACATATGAGTAATACATTTGACAACGCCC	
PDG46_R	TAGCAGCCGGTCGACTTAATGATGATGATGATGATGCGCCACTTTGCGCTTTCG	
PDG80_F	AAGAAGGAGATATACATATGTCAATGCCAACCGATGTG	Overexpression of JamB-M5T (QTY_1/2)
PDG80_R	GTGGTGGTGGTGCTCGAGAGCTAACTTCTTAGCTTCGATC	
PDG86_F	TTTAAGAAGGAGATATACATATGAGTAATACATTTGACAACGCCC	Overexpression of TtuB
PDG86_R	TGGTGGTGGTGGTGCTCGAGCGCCACTTTGCGCTTTCG	
PDG89_R	GTTTCTTTACCAGACTCGAGTTAAGCTAACTTCTTAGCTTCGATC	Co-expression of JamC and JamB-M5T (QTY_1)
PDG97_F	CCTGTACTTCCAATCCCATATGTCAATGCCAACCGATG	Overexpression of MBP-JamB(M5T)
PDG97_R	CGAGTGCGGCCGCAAGCTTAGCTAACTTCTTAGCTTCGATC	
PDG98_F	CCTGTACTTCCAATCCCATATGAGTAATACATTTGACAACGC	Overexpression of MBP-TtuB
PDG98_R	CGAGTGCGGCCGCAAGCTTCGCCACTTTGCGCTTTCG	
PDG109_F	GAAGGAGATATACATATGAGCAACACTTTTGACAATG	Overexpression of TtuB (QTY_2)
PDG109_R	GGTGGTGGTGCTCGAGTGCGACTTTGCGTTTGCG	
PDG113_F	GAAGGAGATATACATATGAAAATCGAAGAAGG	Co-expression of JamC and MBP-JamB-ApoAI*
PDG113_R	CTTTACCAGACTCGAGTCACTGGGTGTTCAATTTTTTAGTATATTC	

PDG114_F	GAAGGAGATATACCATGGGCATGAAAATCGA AGAAGGTAAAC	Co-expression of JamC and MBP-TtuB-ApoAI*
PDG114_R	GCTCGAATTCGGATCCTCACTGGGTGTTCAAT TTTTTAG	

Table 4-2. Plasmids used in this study.

Plasmid	Vector	Description
pXZ23	pET24b	Overexpression of JamA
pXZ24	pCDFDuet	Overexpression of HsPKS1
pXZ27	pCDFDuet	Co-expression of JamA and HsPKS1
pXZ39	pET24b	Overexpression of JamC
pXZ43	pETDuet	Co-expression of JamC and JamB-M5T
pXZ44	pETDuet	Expression of JamC
pXZ64	pET24b	Overexpression of TtuA
pXZ80	pET24b	Overexpression of TtuC
pXZ92	pCDFDuet	Co-expression of TtuA and HsPKS1
pXZ101	pETDuet	Co-expression of TtuC and TtuB
pXZ178	pET24b	Overexpression of MBP-JamB-ApoAI*
pXZ179	pET24b	Overexpression of MBP-TtuB-ApoAI*
pXZ185	pET24b	Overexpression of sfGFP-JamB-ApoAI*
pXZ186	pET24b	Overexpression of sfGFP-TtuB-ApoAI*
pXZ273	pETDuet	Co-expression of JamC and MBP-JamB-ApoAI*
pXZ274	pETDuet	Co-expression of JamC and MBP-TtuB-ApoAI*
pDG12	pJL1	CFPS of JamB (C-His)
pDG19	pETDuet	Expression of JamB-M5T
pDG20	pJL1	CFPS of JamB-sfGFP (C-His)
pDG25	pJL1	CFPS of TtuB (C-His)
pDG50	pET24b	Overexpression of JamB-M5T (QTY_1)
pDG54	pET24b	Overexpression of TtuB
pDG55	pET24b	Overexpression of TtuB (QTY_1)
pDG51	pET24b	Overexpression of JamB-M5T (QTY_2)
pDG59	pETDuet	Co-expression of JamC and JamB-M5T (QTY_1)
pDG65	pET24b	Overexpression of MBP-JamB(M5T)
pDG66	pET24b	Overexpression of MBP-TtuB
pDG76	pET24b	Overexpression of TtuB (QTY_2)
pDG79	pETDuet	Co-expression of JamC and MBP-JamB-ApoAI*
pDG80	pETDuet	Co-expression of JamC and MBP-TtuB-ApoAI*

Construction of plasmids for protein expression in *E. coli*. Individual genes were PCR amplified from previous lab members' plasmids. All primers are listed in **Table 4-1**. Gibson assembly was used to ligate PCR products to linear pET24b vector. Plasmids were isolated using a Zyppy Miniprep Kit (Zymo Research) and confirmed by DNA sequencing (UC Berkeley DNA Sequencing Facility). The plasmid pJL1 was purchased from Addgene. Sequences of JamB(QTY) and TtuB(QTY) in pDG50, pDG51, pDG55 and pDG76 were synthesized by IDT gblocks.

Overexpression and purification of recombinant protein. Expression strains were grown in 1 L of LB supplemented with 50 μg/mL of kanamycin at 37 °C and 220 rpm until an OD_{600} of 0.6. Cultures were then cooled on ice for 10 minutes before induction with 0.2 mM IPTG. Induction of gene expression lasted for 16 h at 16 °C and 200 rpm. The cells were then harvested by centrifugation (6000 x g, 15 min, 4 °C), and supernatant was removed. The pellet was resuspended in 30 mL of lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM imidazole) and lysed with a Q700 sonicator. The insoluble fraction was removed by centrifugation (15,000 rpm, 1 h, 4 °C), and the supernatant was filtered with a 0.45 μM filter before batch binding. Ni-NTA resin (Qiagen) was added to the filtrate at 2 mL/L of cell culture, and samples were allowed to nutate for 1 hour at 4 °C. The protein resin mixture was added to a gravity filter column. The flow through was discarded, the column was then washed with 20-30 mL of wash buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 20 mM imidazole) and tagged protein was eluted in approximately 8 mL of elution buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 250 mM imidazole). Complete elution was determined by Bradford assay. Purified proteins were then concentrated and exchanged into appropriate buffer (50 mM HEPES, pH 7.5, 100 mM NaCl) using Amicon ultra filter units. After two rounds of exchange and concentration, pure protein was removed, and glycerol was added to a final concentration of 8%. Proteins were flash frozen in liquid nitrogen and stored at −80 °C or used immediately for in vitro assays. The presence and purity of enzymes was assessed using SDS-page and concentration was determined using a NanoDrop UV-Vis spectrophotometer (Thermo Scientific).
For the collection of membrane fractions of JamB/TtuB variants, after cell lysis, cell debris was removed by centrifugation (10,000 g, 10 min, 4 °C). The supernatant was further ultracentrifugated (100,000 g, 1 h, 4 °C) to collect the cell membrane. The supernatant was considered supernatant fractions for *in vitro* assays. The membrane fraction was used in the *in vitro* assays immediately.

Enzymatic reactions of standards for compound 8 and 9. A reaction mixture contained 50 mM HEPES (pH 8.0), 2.5 mM $MgCl_2$, 5 mM ATP, 5 mM 5-hexynoic or 5-hexenoic acid, 1 mM CoA, 1 mM Malonyl CoA, and 50 μM HsPKS1. The reaction was initiated by the addition of 50 μM Orf35.[139] The reaction was quenched with the addition of 1% acetic acid in ethyl acetate. Precipitated proteins were removed by centrifugation.[73]

Activity tests of solubilized JamB/TtuB *in vivo*. *E. coli* BAP1 strains containing plasmids of *jamABC*/*ttuABC* and *hspks1* were grown in 0.5 mL LB medium for seeding overnight at 37 °C with 100 μg/ml carbenicillin and 50 μg/ml spectinomycin. The seed cultures were inoculated into 75 mL LB medium to grow until OD_{600} reached 0.6. Cultures were then cooled on ice for 10 minutes before induction with 500 μM IPTG, fed with 1 mM 5-hexenoic acid or 1 mM 9-decenoic acid. Cultures tested in SIMPLEx method were transferred to 25 mL F1 medium with antibiotics, IPTG and substrates when OD_{600} reached 0.6. After 48 h of growth at 18 °C, the culture broth and the cell pellets were separated by centrifugation. Compounds were extracted from cell-free supernatant and cell pellets using ethyl acetate. The solvent was removed by rotary evaporation. The residue was redissolved in 0.5 mL methanol and analyzed by LC/HRMS (20 μL injection). LC/HRMS analysis was performed on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2–98% CH_3CN (v/v) over 20 min in H_2O with 0.1% (v/v) formic acid at a flow rate of 0.5 ml/min was

used. LC-HRMS/MS analysis was performed using the same program described above with untargeted mass/mass and a collision energy of 15 eV.

F1 medium used for *jam*/*ttu* systems. 1L of F1 medium contains 3 g KH_2PO_4, 6.62 g K_2HPO_4, 4 g $(NH_4)_2SO_4$, 150.5 mg $MgSO_4$, 5–20 g glucose, 1.25 mL trace metal solution (optional), 100 μM $Fe(NH_4)_2(SO_4)_2$, and 10 mL 100x vitamin solution (optional). Trace metal solution is composed of 40 mg $ZnCl_2$, 200 mg $FeCl_3 \cdot 6H_2O$, 10 mg $CuCl_2 \cdot 2H_2O$, 10 mg $MnCl_2 \cdot 4H_2O$, 10 mg $Na_2B_4O_7 \cdot 10H_2O$, 10 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ in 1 L of DI water. 100x vitamin solution contains 100 mg/L biotin, 100 mg/L choline chloride, 100 mg/L D-calcium pantothenate, 100 mg/L folic acid, 100 mg/L nicotinamide, 100 mg/L pyridoxal hydrochloride, 10 mg/L riboflavin, 100 mg/L thiamine hydrochloride, 200 mg/L i-inositol, and 8.5 g/L NaCl.[68]

Activity tests of solubilized JamB/TtuB *in vitro*. Coupled enzymatic reactions for JamB variants tests were performed at room temperature for 5 hours in 50 μL of 50 mM HEPES (pH 8.0) containing 2.5 mM $MgCl_2$, 5 mM ATP, 1 mM TCEP, 5 mM 5-hexenoic acid, 1 mM NADH, 1 mM NADPH, 1 mM $Fe(NH_4)_2(SO_4)_2$, 7.5 μg Fx, 25 mU FxR, 1 mM PMS, 1 mM ascorbic acid, 300 U catalase, 1 mM Mal-CoA, 10 μM JamA, 50 μM JamC, 50 μM HsPKS1 and 10 μM JamB variants, or JamB membrane fractions (resuspended in exchange buffer, add till the volume reached 50 μL), or JamB supernatant fractions (add till the volume reached 50 μL). Similarly, coupled enzymatic reactions for TtuB variants tests were performed at room temperature for 5 hours in 50 μL of 50 mM HEPES (pH 8.0) containing 2.5 mM $MgCl_2$, 5 mM ATP, 1 mM TCEP, 5 mM 9-decenoic acid, 1 mM NADH, 1 mM NADPH, 1 mM $Fe(NH_4)_2(SO_4)_2$, 7.5 μg Fx, 25 mU FxR, 1 mM PMS, 1 mM ascorbic acid, 300 U catalase, 1 mM Mal-CoA, 10 μM TtuA, 50 μM TtuC, 50 μM HsPKS1 and 10 μM TtuB variants, or TtuB membrane fractions (resuspended in exchange buffer, add till the volume reached 50 μL), or TtuB supernatant fractions (add till the volume reached 50 μL). The reaction was quenched and extracted by ethyl acetate with 1% acetic acid. The organic solvent dried under N_2, then the extracts were redissolved in 100 μL MeOH and injected 20 μL for LC/HRMS analysis.

Cell-free protein synthesis of JamB/TtuB. Making CFPS reagents, CFPS and measurement of fluorescence were according to the published protocol.[140] The second and third round of CFPS was based on the manual of NEBExpress® Cell-free E. coli Protein Synthesis System.

Chapter 5. Modular Assembly Line Engineering of Terminal Alkyne Tagged Natural Products

5.1 Introduction

The 2022 Nobel Prize in chemistry was awarded to scientists Carolyn R. Bertozzi, Morten P. Meldal and K. Barry Sharpless in recognition of their groundbreaking contributions to the development of click chemistry and bioorthogonal chemistry. The terminal alkyne functionality, due to its chemical stability in biological environments, its rarity in biology, and its selectivity in bioorthogonal reactions, has emerged as one of the most powerful tools in chemical biology. Terminal alkyne tags serve as versatile chemical handles for bioorthogonal chemistry, enabling the selective labeling and tracking of natural products in complex biological systems.[141] This facilitates the elucidation of their biosynthetic pathways, biological targets, and mechanisms of action. In addition, terminal alkyne tagged natural products, coupled with azido probes or spectroscopic methods, offer a robust tool for high throughput screening assays, enabling visualization and quantification of biological activities.[119,142] Furthermore, terminal alkyne tags facilitate the chemical synthesis and modification of natural products, promoting the generation of structurally diverse analogs for structure-activity relationship studies.[143] Therefore, the terminal alkyne tags amplifies our capacity to explore, manipulate, and exploit the biological activities and therapeutic potential of natural products across a myriad of research fields and applications.

Tagging natural products with the terminal alkyne functionality can be achieved through multiple ways including total synthesis, semi-synthesis, precursor-directed biosynthesis (PDB), and *de novo* biosynthesis.[29,73,142,144] Major challenges for total synthesis and semi-synthesis stem from the intricate structural complexity and restricted availability of natural products. PDB relies on feeding terminal alkyne-containing precursors, leading to the coexistence of alkyne tags in both starting materials and final products, which renders this method incompatible with *in situ* bioorthogonal chemical transformations. Therefore, *de novo* biosynthesis emerges as a standout approach, as it can produce terminal alkyne-tagged natural products while uniquely enabling the *in-situ* utilization of alkyne tags. The discovery of the portable tri-gene cassettes, *jamABC* and *ttuABC*, has paved the way for the on-demand *in situ* generation and incorporation of terminal alkynes into various molecular scaffolds. Previously, our lab successfully achieved the *de novo* biosynthesis of terminal alkyne tagged polyketides with type III PKSs and type I PKSs.[73,77,78]

Type III PKSs, as simple homodimers, catalyze the whole series of decarboxylation, condensation, and cyclization reactions with a single active site. Type III PKSs perform C–C bond forming reactions by iterative Claisen-type condensation of CoA thioesters and cyclization of the poly-β-keto intermediates, to produce pharmaceutically and biologically important aromatic polyketides, such as chalcone, stilbene, and curcumin.[145,146] Additionally, Type III PKSs often exhibit remarkable substrate tolerance toward different acyl-CoA starter units, representing an excellent platform in generating novel polyketides and making starter unit engineering feasible. Considering the substrate specificity of JamB and TtuB, we selected HsPKS1 for *de novo* biosynthesis of terminal alkyne tagged polyketides. Co-expression of genes encoding *jamABC*/*ttuABC* and *hspks1* in *E. coli* BAP1 strains with the respective feeding of 5-hexenoic acid or 9-decenoic acid yielded the successful production of compound **6** and **8**. This marked a

significant milestone, demonstrating the feasibility of incorporating type III PKSs in biosynthetic pathway engineering to generate terminal alkyne tagged natural products.

In contrast, type I PKSs are multifunctional enzymes containing linearly arranged and covalently fused domains with defined functions separated by short-space regions. The distinct domains work cooperatively to catalyze the carbon chain elongation and functional group modification, and act as an assembly line to select and incorporate building blocks into polyketide scaffolds.[21] The mega-synthases have garnered considerable attention and have been the subject of extensive engineering endeavors owing to their modular structures, which have captivated researchers with the prospect of synthesizing customized molecules on demand.[147] The *de novo* biosynthesis of terminal alkyne tagged polyketides can be accomplished by leveraging JamABC to provide a building monomer through starter unit engineering and extender unit engineering.[73,78]

To explore a generalized method of starter unit engineering, Zhang lab employed two well-studied type I PKSs, LipPKS1 and DEBSM6. LipPKS1 is the first module in lipomycin biosynthesis that natively uses an isobutyl starter unit presented by a loading ACP.[148] DEBSM6 is the last PKS module from the erythromycin biosynthetic pathway.[149] Considering the known critical role of the cognate ACP (JamC) in the alkyne biosynthetic machinery,[118] the recognition of JamC by PKSs is expected to play a critical role in alkyne-tagged polyketide synthesis and therefore is the focus of the engineering. The researchers created a truncated version of PKSs to facilitate the alternative starter unit incorporation. These engineered PKSs (termed LipPKS1* and DEBSM6*) without the loading domains were expected to result in JamC acting in *trans* to selectively load and extend JamC-tethered acyl chains. JamB activity for alkyne biosynthesis was assessed *in vivo* due to the difficulty of obtaining active and purified membrane proteins. Two engineering strategies, docking domain and site-directed mutagenesis, have been demonstrated to enhance JamC-PKS recognition (**Figure 5-1**).

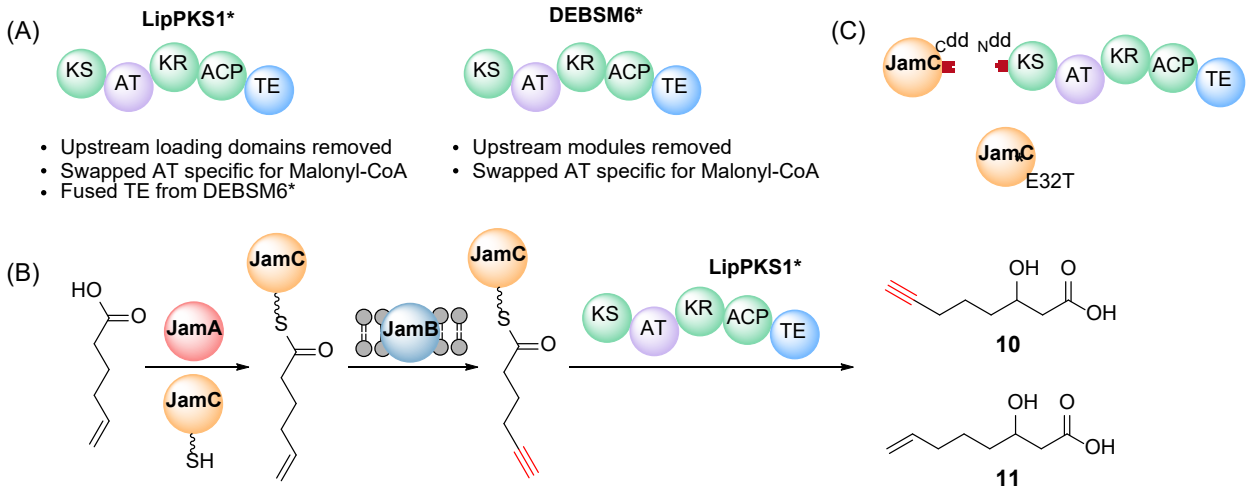


Figure 5-1. Type I PKS starter unit engineering for terminal alkyne tagged polyketide biosynthesis. **(A)** Engineered LipPKS1* and DEBSM6* used as representative type I PKSs. **(B)** Schematic of biosynthetic pathway. **(C)** Two strategies in type I PKS starter unit engineering: docking domain and site-directed mutagenesis.

Docking domains, often found on the C terminus of ACPs (dd^{ACP}) and the N terminus of KSs (dd^{KS}), have been shown to be important for protein-protein interactions in PKSs.[150] Zhang lab fused known docking domains to the C terminus of JamC and the N terminus of the LipPKS1* /DEBSM6* KS domains to improve protein recognition. *In vitro* product formation assays

demonstrated the success of this strategy in generating product **10** with enhancement of product titer of 3-fold for LipPKS1* and 40-fold for DEBSM6*. In addition to docking domain engineering, a less-intensive engineering strategy, site-directed mutagenesis, to improve JamC-PKS interaction was studied. From the well-studied DEBS system, it has been shown that direct ACP-KS protein-protein interactions during translocation are selective, and key residues within helix I of ACP have been identified to contribute to chain translocation specificity.[41,151] Zhang lab identified the corresponding residue in JamC (E32) that may play an important role in ACP-KS interactions through sequence alignments and structural modeling. To mimic the native upstream ACP, they chose the mutations E32T for LipPKS1* and E32H for DEBSM6* based on alignments to the respective ACPs found upstream in the native systems. *In vitro* product formation assays showed that the production of **10** increased approximately 7-fold with LipPKS1* and 2-fold with DEBSM6*. It demonstrated the effectiveness of this strategy in improving the production of alkyne-tagged polyketides *in vitro*, most likely due to an improved JamC communication with modular PKSs.

The extender unit engineering is an alternative strategy for incorporating an alkynoic extender unit into the molecular scaffolds of PKs besides starter unit engineering. Zhang lab applied the JamABC cassette to the biosynthetic pathway for antimycin-type depsipeptides.[73] It has been shown that an alkynoic extender unit can be recognized by the promiscuous AntE and AntD-AT to synthesize an alkyne-labeled antimycin by PDB in *Streptomyces* species.[152] To generate terminal alkyne tagged antimycin analogs by *de novo* biosynthesis, they co-expressed *antCDEFGM* and *jamABC* in the *E. coli* BAP1 strain and achieved the goal.

With successful engineering interventions between JamC and PKSs, our aim is to broaden the scope of applications to generate terminal alkyne-tagged lipopeptides by engineering the interaction between JamC and NRPSs. Like type I PKSs, NRPSs are modular megasynthases. The modular architecture offers unparalleled flexibility and versatility, making it an attractive platform for the rational design and engineering of novel bioactive compounds. However, the investigation into the interplay between ACPs and NRPSs remains underexplored, representing a gap we aim to address in our research. Lipopeptides have emerged as important antifungal and antibacterial agents widely used in food safety and therapeutics.[153] The rise of drug-resistant bacteria has underscored the need for new antimicrobial molecules like lipopeptides. Daptomycin, as the first cyclic lipopeptide antibiotic approved by Food and Drug Administration (FDA), shows a broad range of activity against all gram-positive bacteria including methicillin, vancomycin, and linezolid resistant organisms. *Bacillus* species are efficient producers of lipopeptides, including surfactin, iturin, and fengycin, which exhibit potent antagonistic activities against plant pathogens. These compounds offer promising alternatives to conventional antibiotics and antifungal treatments. Additionally, lipopeptides from *Streptomyces* species and synthetic derivatives serve as pharmaceuticals and vaccine adjuvants with various applications. Overall, lipopeptides present versatile solutions to combat drug-resistance and address global challenges in medicine, industry, and agriculture. Incorporating terminal alkyne tags into lipopeptides not only enables in-depth studies of biological activities, modes of action and potential therapeutic benefits, but also facilitates the production of novel derivatives and potentially enhanced efficacy. Therefore, we aim to employ the JamABC/TtuABC biosynthetic machinery and assembly line engineering strategies to generate terminal alkyne tagged lipopeptides.

The two model NRPS systems we chose were the EpnG in the biosynthesis of eponemycin and the MmaA in the biosynthesis of isonitrile lipopeptides (INLPs).[154,155] Firstly, epoxomicin and epoxomicin are proteasome inhibitors with α',β'-epoxyketone pharmacophores, which allow

binding to the N-terminal threonine in catalytic β-subunits of the proteasome with extraordinary specificity.[156] Zhang lab reported the identification of the minimal set of enzymes from the eponemycin and epoxomicin BGC necessary for the biosynthesis of novel metabolites containing a terminal epoxyketone pharmacophore in *E. coli* (**Figure 5-2**). EpnG serves as the initial NRPS module containing two sets of CAT domains for chain elongation. The first C domain in EpnG is relatively promiscuous towards fatty acyl chain length and it is demonstrated that acyl CoA substrates can be accepted *in vivo*. Then the A domains activate and load amino acids, specifically serine and leucine, in sequential order to form two amide bonds, thereby elongating the molecular chain tethered to the carrier proteins. By incorporating terminal alkyne tags, researchers can unlock new avenues for exploring the mechanisms underlying the pharmacological effects and therapeutic potential of the epoxyketone moiety.

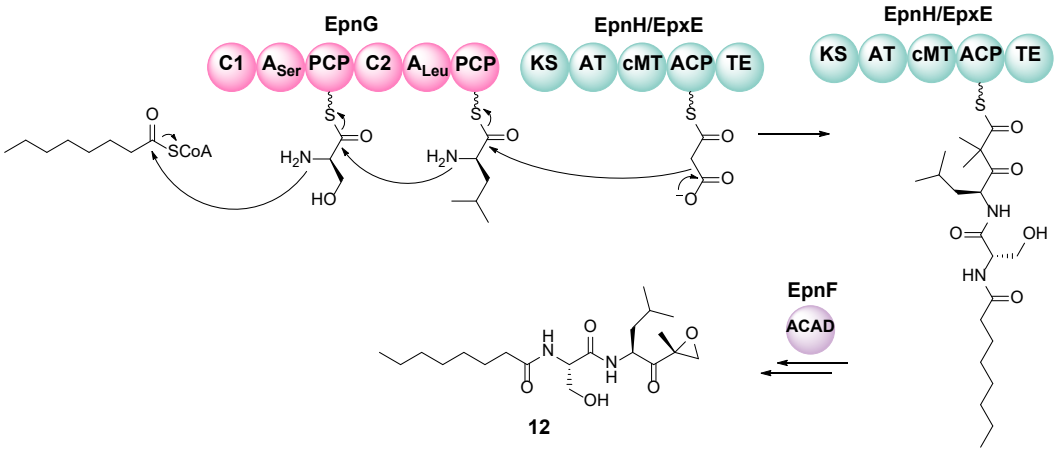


Figure 5-2. Proposed biosynthesis of a peptidyl epoxyketone produced in *E. coli*. CMT stands for C-methyltransferase. ACAD stands for acyl-CoA dehydrogenase.

Secondly, MmaA is an NRPS module identified from *Mycobacterium marinum*. Homologous genes are conserved in many Actinobacteria species, including *M. tuberculosis*. Although the native BGC metabolites remain unknown, co-expression of *mmaABCDE* in *E. coli* led to the production of INLPs (**Figure 5-3**). INLPs are known to be related to the virulence of pathogenic mycobacteria by mediating metal transport.[157] In *mma* pathway, MmaA, acting as the single-module NRPS, promotes the condensation of isonitrile-containing substrates to lysine, followed by reductive release to form INLPs. This system offers an excellent platform for investigating the effects of biosynthetic engineering, owing to its inherent straightforwardness. There are two primary motivations for tagging a terminal alkyne to INLPs within this system. First, it is significant tag a terminal alkyne moiety to INLPs for downstream applications and functional studies. Second, the simplicity of this system makes it an ideal model for testing various engineering strategies and assessing their impact on biosynthesis.

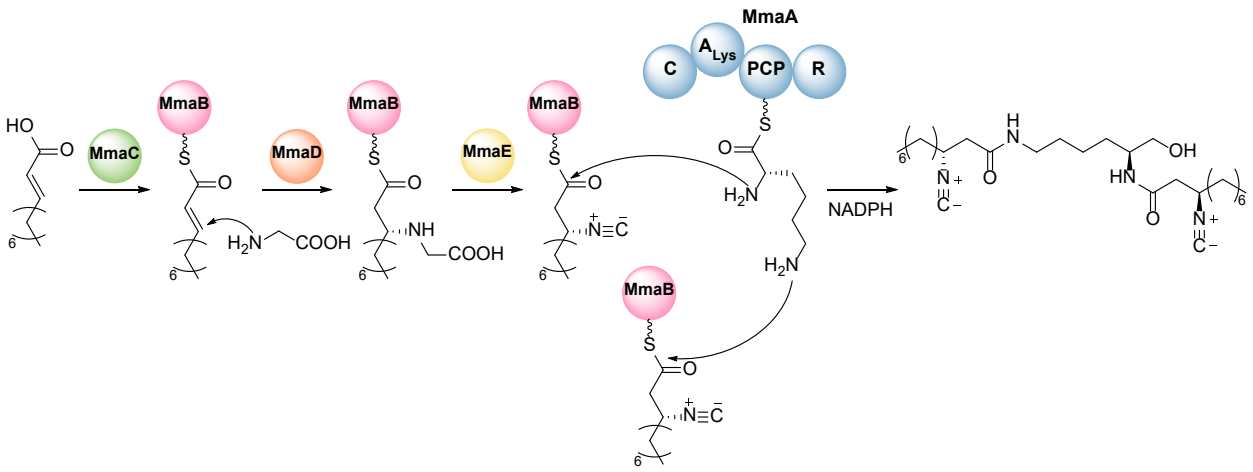


Figure 5-3. Proposed biosynthesis of an isonitrile lipopeptide produced in *E. coli* with *mma* BGC.

In this chapter, we harnessed the *de novo* biosynthesis machinery of terminal alkyne, TtuABC, and applied it to two model NRPS systems, EpnG and MmaA, with the aim of generating terminal alkyne tagged lipopeptides. We investigated the interactions between ACPs and NRPSs by engineering the biosynthetic pathway *in vivo* and *in vitro*. Our engineering efforts encompassed a range of approaches, including combinatorial biosynthesis, site-directed mutagenesis and a novel transacylase between carrier proteins.

5.2 Results

5.2.1 Biosynthetic engineering *in vivo* in EpnG system

In collaboration with my undergraduate mentee, Yuxin Xie, we investigated the possibility of generating terminal alkyne tagged lipopeptides with the biosynthetic pathway of eponemycin (*epn*) and epoxomicin (*epx*). From previous studies, co-expression of *epnG*, *epxE* and *epnF* in E. coli BAP1 strains yielded the production of compound **12** with an epoxyketone pharmacophore (**Figure 5-2**). Inspired by the starter unit engineering in type I and type III PKSs, we co-expressed *ttuABC* together with *epnG*, *epxE* and *epnF* in *E. coli* BAP1 strains in LB medium with the supplementation of 0.5 mM of IPTG, 1mM of serine, 1mM of leucine and 1mM of 9-decenoic acid (**Figure 5-4**). Although the first condensation domain in EpnG prefers C_8 acyl chain length, it has broad substrate scope, making it possible to accept C_{10} fatty acyl chain after the modification of TtuABC. Therefore, the challenge here lay in whether EpnG was able to recognize TtuC efficiently. To study this problem, we compared the performance in producing compound **13** of individually expressed TtuC and EpnG with that of fused TtuC and EpnG.

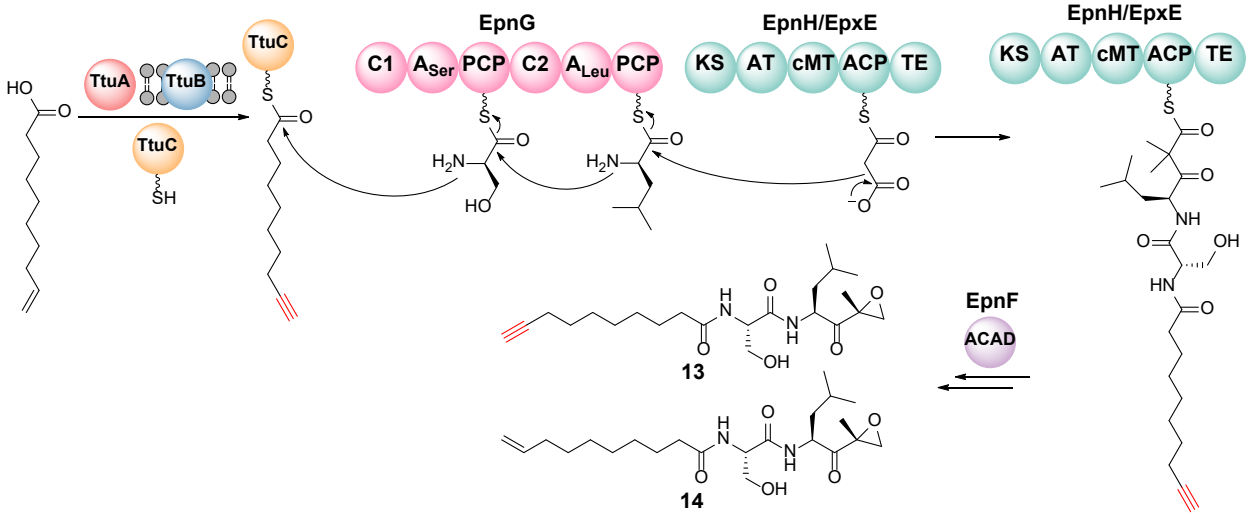


Figure 5-4. Proposed biosynthesis of terminal alkyne tagged epoxyketone metabolites in *E. coli*.

The constructs we used included pCDFDuet vector bearing *ttuA* and *epxE*, pCOLADuet vector bearing *epnF*, and pETDuet vector bearing *ttuC*, *epnG* and *ttuB*. Each gene was regulated under a T7 promoter except for *ttuC* and *epnG*. In one strain, *ttuc* and *epnG* were individually expressed under one T7 promoter, while in the other, *ttuC* was fused to *epnG*. The growth conditions of both strains were identical. We harvested cultures after 48-hour growth at 18 ºC and extracted supernatant and cell pellets separately with ethyl acetate. The extracts were then analyzed by LC/HRMS. Encouragingly, we observed the production of compound **13** in the supernatant of both strains, albeit at low abundance (10^3). Fusing TtuC to EpnG did not increase the production of terminal alkyne tagged epoxyketone compound. The production of compound **14** exhibited a 100-fold higher compared to that of compound **13**, suggesting robust functionality of TtuAC and enzymes in *epn*/*epx* pathway. Remarkably, extracts from both strains revealed a notable presence (10^5) of 9-decynoic acid, underscoring the efficiency of TtuB in catalyzing terminal alkyne formation. However, most alkynes failed to enter *epn*/*epx* pathway. Instead, they got hydrolyzed off from TtuC. The comparison of the quantity of 9-decenoic acid, 9-decynoic acid, compound **13** and compound **14** is shown in **Figure 5-5**, suggesting two potential scenarios. Firstly, it is possible that there exists an endogenous thioesterase enzyme in *E. coli* responsible for hydrolyzing 9-decynoic acid from TtuC. Secondly, it is conceivable that EpnG exhibits poor recognition of TtuC. Previous studies have demonstrated that even in the absence of external feeding, compound **12** can still be produced, as EpnG has the capability to accept substrates from CoA thioesters. One plausible pathway was thus proposed. EpnG failed to recognize TtuC tethered substrates directly. Upon the production of 9-decynoyl-TtuC, 9-decynoic acid accumulated due to hydrolysis from TtuC. Some 9-decynoic acid was incorporated into the *E. coli* fatty acid biosynthesis pathway to generate 9-decynoyl-CoA, which was then recognized by EpnG. The complexity in the pathway resulted in the trace production of compound **13**. Meanwhile, the excess quantity of fed 9-decenoic acid could be utilized by *E. coli* to produce 9-decenoyl CoA, which was also accepted by EpnG to produce a large quantity of compound **14**. In alignment with this hypothesis, we observed the production of various fatty acyl chains incorporated epoxyketone. Therefore, fusing TtuC and EpnG together did not enhance the interaction between them, instead this engineering approach decreased the recognition between TtuAB and TtuC, leading to less production of 9-decynoic acid. Notably, the semi-quantitative analysis was based on HRMS-EIC with the assumption that these

four molecules behave similarly in mass spectrometry. However, we observed limitations in detecting medium and long-chain fatty acids under positive mode in our MS. Thus, we actually underestimated the quantity of 9-decenoic acid and 9-decynoic acid.

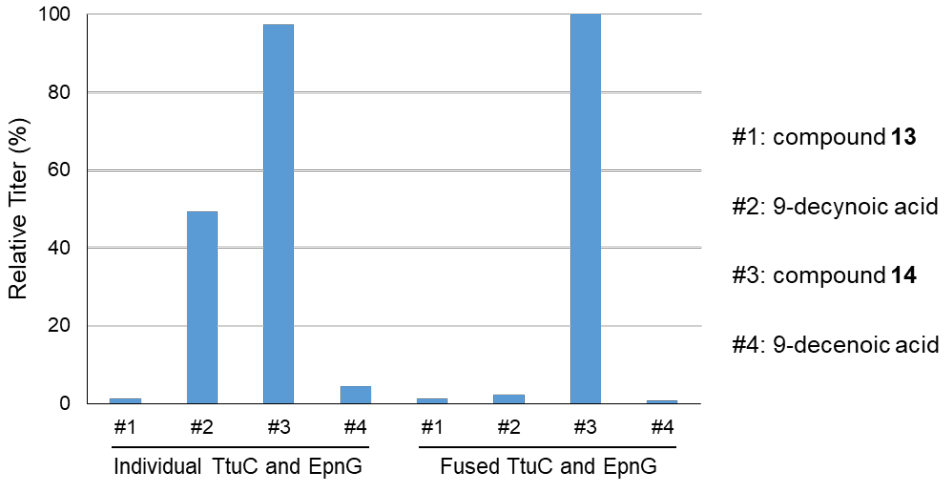
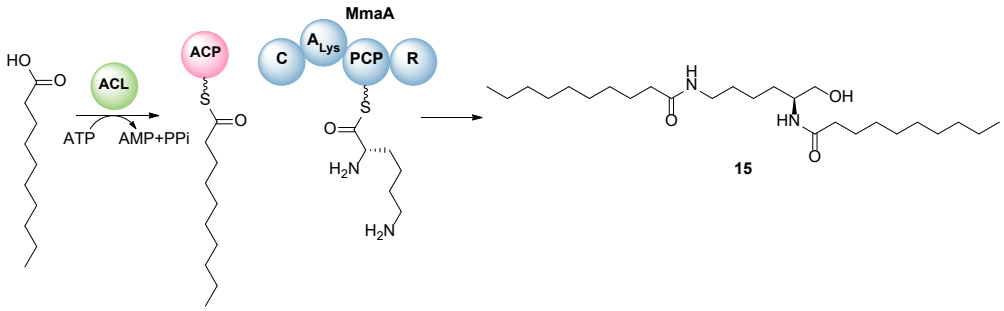


Figure 5-5. Titer comparison of compounds in extracts from supernatant of *E. coli* cultures with the co-expression of *ttuABC*, *epnG*, *epxE* and *epnF*, measured by LC/HRMS. Compound **15** only existed in supernatant. The vast majority of other compounds existed in supernatant, with trace in cell pellets.

5.2.2 Biosynthetic engineering *in vitro* in MmaA system

Drawing from the prior heterologous expression of the *mma* pathway in *E. coli*, we have established that C_{10} fatty acyl chain is a viable substrate for the C domain of MmaA. Therefore, the acyl substrate of TtuABC is compatible with that of MmaA, making it feasible for starter unit engineering. We aimed to investigate the interaction between TtuC and MmaA *in vitro* (**Figure 5-6**), considering that MmaA was a single-module NRPS with an existing releasing domain. More importantly, we had access to its homologous upstream ACP, Rv0100 and activating enzyme, Rv0099, from the INLPs biosynthetic pathway in *M. tuberculosis*. We conducted enzymatic assays involving decanoic acid, TtuAC or Rv0099, Rv0100 and MmaA, validating superior recognition between Rv0100 and MmaA compared to TtuC and MmaA (**Figure 5-7**). By comparing trace **i** and **ii**, using non-cognate ACP decreased the production of compound **15** by approximately 25-fold. Excluding TtuC completely abolished the production of compound **15**, indicating MmaA failed to recognize AMP-activated substrates and the presence of carrier proteins was thus required (**Figure 5-7**, trace **i**, **ii**, **viii**).



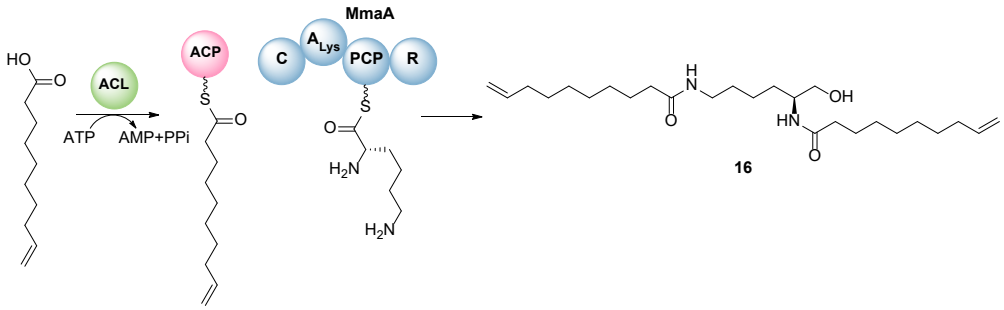
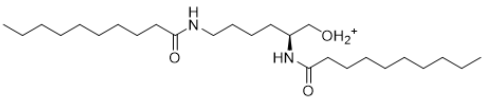


Figure 5-6. Biosynthesis of lipopeptides with MmaA *in vitro*. ACL stands for acyl-CoA ligase, representing TtuA and Rv0099. ACP represents TtuC and Rv0100, respectively.

The first method we employed was site-directed mutagenesis of TtuC. Mutating TtuC without perturbation to the large mega-synthase would make this strategy more easily adaptable to different systems. Nevertheless, the absence of standardized guidelines delineating the pivotal residues governing the interaction between ACPs and NRPSs presents a hurdle to this engineering approach. Fortunately, MmaA exhibited robust recognition of Rv0100, rendering Rv0100 an apt model for investigation and enabling us to engineer TtuC to mimic its behavior. We thus generated a TtuC mutagenesis list of P40D in loop I, S49V, A50R, A56K, S59G, E60V in helix II, and R72A in helix III based on structural modelling and tentative protein-protein docking. These seven TtuC mutants (10 kDa) were cloned, overexpressed, and purified from BAP1 with a similar yield, approximately 10 mg/L, to the wild-type protein (**Appendix M**). Enzymatic assays were performed to compare the activity of WT TtuC and TtuC mutants. Contrary to our expectation, mutants A50R, S59G and R72A showed decreased production of compound **15** while mutants P40D, S49V, A56K and E60V completely abolished the production of compound **15** (**Figure 5-7**). These observations might stem from reduced recognition either between TtuA and TtuC or between TtuC and MmaA, underscoring the nuanced nature of ACP recognition by other biosynthetic enzymes.



m/z: 441.4051 (100.0%), 442.4085 (28.1%), 443.4118 (3.8%)

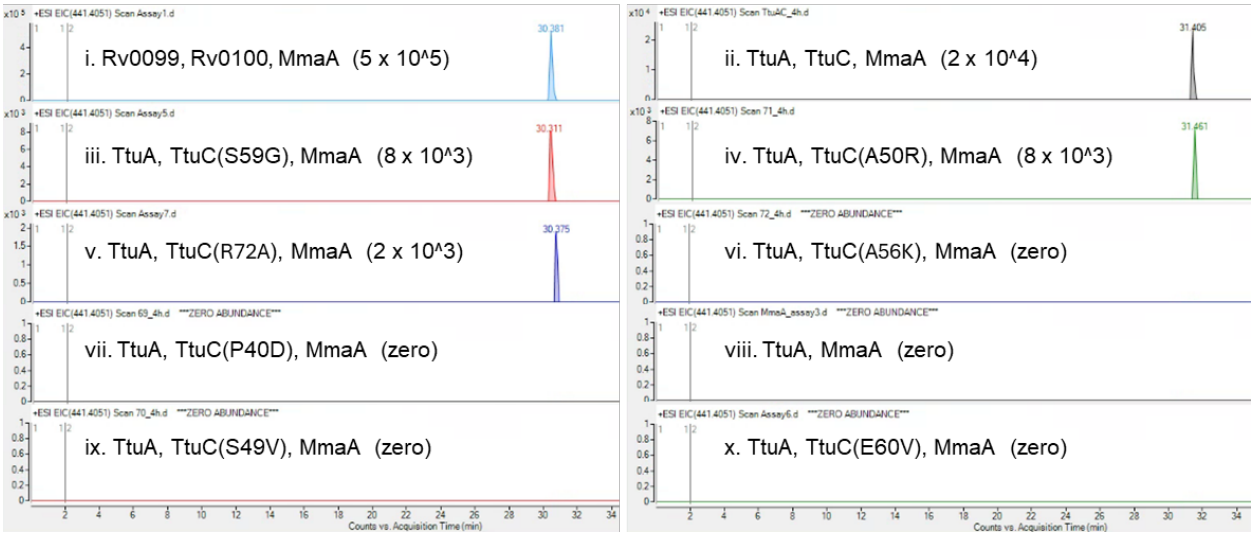


Figure 5-7. Enzymatic assay results of compound **15** production, measured by LC/HRMS. MS/MS confirmation of compound **15** is listed in **Appendix N**. The shift of retention time was due to the maintenance of LC part.

The second method we adopted involved a novel transacylase between carrier proteins. Given the essential role of TtuABC in terminal alkyne formation and the favorable recognition of Rv0100 by MmaA, we intended to employ a transferase to shuttle the acyl chain from TtuC to Rv0100. This approach enables the incorporation of the alkyne-containing starter unit into MmaA, presenting a versatile method applicable to diverse assembly line systems. Non-elongating KS (KS^0) domains, lacking the conserved His residue required for chain elongation, are proposed to transfer biosynthetic intermediates between ACPs.[158] The Challis lab reported a KS^0 domain of Bamb_5919 involved in enacyloxin biosynthesis with extensive biochemical characterization to delineate its function. The KS^0 domain exhibits the ability to transfer the fully assembled polyketide chain from the upstream ACP to a downstream PCP, facilitating the enacyloxin biosynthesis.[159] During the characterization of the KS^0 domain, they used simplified acyl chain substrates, suggesting the KS^0 domain possesses relaxed substrate specificity. This finding served as inspiration for our utilization of the KS^0 domain in the TtuC and MmaA system. Thanks to the Challis lab for generously providing the plasmid encoding the KS^0 domain. After we obtained the purified KS^0 domain from *E. coli* BL21(DE3)* strain, we conducted enzymatic assays involving 9-decenoic acid, TtuAC or Rv0099/Rv0100, KS^0 and MmaA. The negative control omitting KS^0 was performed to demonstrate the role of the KS^0 domain (**Figure 5-8**). In the presence of both TtuC and Rv0100, they competed for interactions with MmaA, resulting in minimal production of compound **16**. However, adding the KS^0 domain to the system significantly increased the production of compound **16** by 10-fold, suggesting that KS^0 domain facilitates the transfer of acyl substrates between these two ACPs. We conducted additional negative controls to ensure thoroughness and accuracy. It is evident that TtuA cannot recognize Rv0100 based on trace **ii** in **Appendix O**. The results of trace **iv** and **vii** eliminated the possibility that KS^0 could complement the functions of ACPs or TtuA. Therefore, the reaction process in the presence of all enzymes can be proposed as follows. TtuA activated and loaded 9-decenoic acid onto TtuC, followed by the KS^0 domain transferring the 9-decenoyl acyl chain from TtuC to Rv0100, which is well recognized by MmaA for the lipopeptide production. Direct evidence showing the acyl chain transferred from TtuC to Rv0100 will be collected through ACP native MS. The only limitation observed thus far is that the adoption of KS^0 did not substantially increase the production of compound **16** compared to the TtuAC and MmaA system alone (trace **viii**), possibly because TtuC is still recognized by MmaA to some extent. Additionally, it is possible that the recognition of KS^0 to TtuC and Rv0100 is not ideal. The transfer of acyl chains is not limited to one direction; it is also possible for transfer from Rv0100 to TtuC, which could be another contributing factor. However, the application of KS^0 in TtuC, Rv0100 and MmaA system demonstrated the mediating role played by KS^0 between ACPs, which is a versatile method for other multi-ACP systems. Furthermore. The investigations on the generation of terminal alkyne tagged lipopeptides by MmaA system will be carried out *in vivo* owing to the insolubility of TtuB.

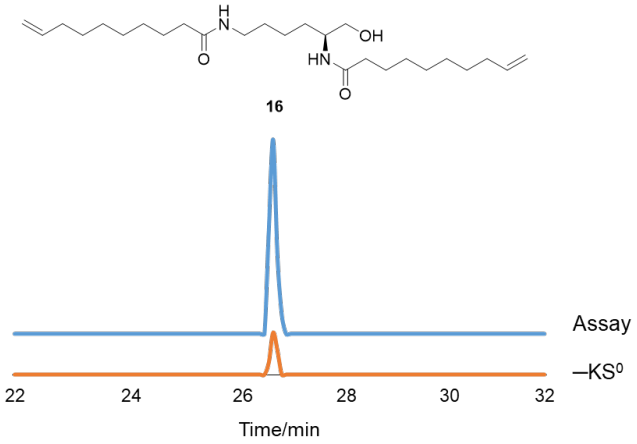


Figure 5-8. Enzymatic assay results of compound **16** production, measured by LC/HRMS. MS/MS confirmation of compound **16** is listed in **Appendix N**.

5.3 Discussion

The results of biosynthetic engineering in EpnG system indicated simple fusion of non-cognate ACPs and NRPSs is not an ideal approach to promote recognition between them. More importantly, the inclination towards CoA substrates led to the complicated metabolite profiles in vivo, thereby diminishing the feasibility of integrating TtuABC as a starter unit in engineering efforts. This highlights the importance of selecting appropriate NRPS systems. Nevertheless, for *in vitro* biochemical studies involving TtuC and EpnG, we plan to incorporate the KS^0 domain of Bamb_5919 into the system. The groundwork for this, including the fusion of EpnG with a versatile TE domain and the identification of a suitably recognized ACP for EpnG, has been successfully established.

Non-elongating KS (KS^0) domains are commonly found in *trans*-AT PKSs. Due to the absence of the conserved His residue in the HTGTG motif which is essential for decarboxylative condensation, KS^0 domains lack the ability to elongate intermediates along the assembly line.[158] The KS^0 domain of Bamb_5919 has an Ala residue where regular KS domains have a His residue, and it has been demonstrated to possess a Cys residue at the active site for transacylation.[159] The KS^0 domain acted as a mediator between two ACPs in the MmaA system, suggesting its versatility as a method for integrating different ACPs with modular megasynthases. To broaden the applicability of this method, additional engineering of the KS^0 domain to accommodate various carrier proteins in biosynthetic pathways can be explored, potentially enabling the production of a wider array of valuable compounds.

To effectively incorporate the terminal alkyne tag generated by JamABC/TtuABC into downstream NRPSs, it is crucial to augment the recognition between JamC/TtuC and the condensation domain in NRPSs. While research on the interactions between ACPs and C domains is limited, insights can be gleaned from studying the protein-protein interface between PCPs and C domains. Analysis of multiple C domain crystal structures reveals a distinct division into two halves, commonly referred to as the N-terminal lobe and C-terminal lobe.[160,161] The role of C domains as a secondary checkpoint has been widely discussed. It is proposed that before catalyzing peptide bond formation, the C domain must first bind the correct substrates in its active site. Moreover, it is essential for the C domain to establish specific protein-protein interactions with both the appropriate donor and acceptor PCPs. Understanding the intricacies of protein-protein

recognition between the C domain and both donor and acceptor PCPs is indispensable, particularly in the context of engineering C domains with non-cognate biosynthetic systems.[162] The crystal structure of the 144-kilodalton *Bacillus subtilis* termination module SrfA-C revealed the PCP-C interactions. The A and C domains of SrfA-C associate closely to form a catalytic platform, with their active sites on the same side of the platform. The PCP is flexibly tethered to this platform and thus can move with its substrate-loaded PPant arm between the active site of the A domain and the donor side of the C domain.[163] The structural analysis of the SrfA-C acceptor PCP-C domain unveiled a protein-protein interface involving the C domain N-terminal lobe and C-terminal lobe with the PCP helix 2 and helix 3, respectively. These interactions were predominantly hydrophobic, with minimal involvement of potential hydrogen bonding interactions. Similarly, the complex structure of donor PCP-C domain from linear gramicidin synthetase subunit A revealed a protein-protein interface mainly dependent on hydrophobic interactions.[164] These interactions are located at the PCP loop 1, helix 2, and helix 3 regions and the C domain C-terminal lobe. The PCP loop 1 contributes an electrostatic interaction with C domain as well. Overall, hydrophobic interactions predominate in the interface between PCPs and C domains, regardless of whether C domains interact with donor or acceptor PCPs.

In summary, we utilized two NRPS systems, EpnG and MmaA, to assess the efficacy of starter unit engineering in *in situ* producing terminal alkyne tagged lipopeptides *in vivo* and *in vitro*. Our engineering endeavors focused on enhancing the recognition of ACPs by NRPSs, enabling EpnG or MmaA to efficiently accept acyl chains containing alkyne moieties for condensation and pass along the assembly line. It turned out that simple fusion of TtuC to EpnG did not enhance recognition; instead, it decreased the recognition from TtuA/TtuB. However, the incorporation of the KS^0 domain in MmaA system facilitated the transfer of acyl chains from TtuC to Rv0100, thus promoting the formation of terminal alkyne tagged lipopeptides. The innovative engineering strategy involving the KS^0 domain provides valuable insights into the dynamic interplay between ACPs and NRPSs, unlocking new avenues for the production of tailored molecules with enhanced functionalities.

5.4 Materials and Methods

Chemicals, reagents and strains. Decanoic acid was purchased from TCI Chemicals. Other chemicals were purchased from Fisher Scientific or Sigma-Aldrich. Phusion High-Fidelity PCR Master Mix (NEB) was used for PCR reactions. NEBuilder HiFi DNA Assembly Master Mix was used for DNA construct cloning. Restriction enzymes were purchased from Thermo Fisher Scientific. Bacterial strains involved in this study include *E. coli* XL1 Blue for DNA cloning and amplification, *E. coli* BL21 (DE3)* for recombinant protein overexpression, and *E.coli* BAP1 for recombinant protein overexpression and compound heterologous expression.

Table 5-1. Oligonucleotides used in this study.

Primer	Sequence (5' → 3')	Notes
PDG102_F	GATACGGATTTCCACCAGTTCGGGCTGG	Overexpression of TtuC(P40D)
PDG102_R	GGAAATCCGTATCAAAATCAATCGCACTTTCGC	
PDG103_F	GATTCGGTTGCTGTTGTGGGTTTAGTGGCAG	Overexpression of TtuC(S49V)
PDG103_R	GCAACCGAATCCAGCCCGAACTGGTGG	
PDG104_F	CGTCTCGTGTTGTGGGTTTAGTGGCAGATC	Overexpression of TtuC(A50R)
PDG104_R	CAACACGAGACGAATCCAGCCCGAACTGG	

PDG105_F	GTGAAAGATCTTAGTGAGTGGAGTGGAGTG	Overexpression of TtuC(A56K)
PDG105_R	GATCTTTCACTAAACCCACAACAGCAGACG	
PDG106_F	GATCTTGGTGAGTGGAGTGGAGTGGCGATAG	Overexpression of TtuC(S59G)
PDG106_R	CACTCACCAAGATCTGCCACTAAACCCACAAC	
PDG107_F	CTTAGTGTGTGGAGTGGAGTGGCGATAGG	Overexpression of TtuC(E60V)
PDG107_R	CTCCACACACTAAGATCTGCCACTAAACCC	
PDG108_F	CCATCGCCAAAAATAATTCTATCCATGCGC	Overexpression of TtuC(R72A)
PDG108_R	TTTTTGGCGATGGATTTAATACCTATCGCCAC	
PDG111_F	CTTTAAGAAGGAGATATACATATGCGGGACCGAATCCTCGCC	Overexpression of Rv0100
PDG111_R	GTGGTGGTGGTGGTGCTCGAGACCGAACTCGGTGCACAC	
PDG112_F	CTTTAAGAAGGAGATATACATATGACCGCGCCCGAAATCGG	Overexpression of MmaA
PDG112_R	GTGGTGGTGGTGGTGCTCGAGGGCCAGTGCGGGTCTTGGC	

Table 5-2. Plasmids used in this study.

Plasmid	Vector	Description
pXZ64	pET24b	Overexpression of TtuA
pXZ80	pET24b	Overexpression of TtuC
pXZ255	pCDFDuet	Co-expression of TtuA and EpxE
pXZ256	pCOLADuet	Expression of EpnF
pXZ258	pETDuet	Co-expression of TtuC_EpnG (individual proteins) and TtuB
pXZ260	pETDuet	Co-expression of TtuC-EpnG fusion and TtuB
pDG69	pET24b	Overexpression of TtuC(P40D)
pDG70	pET24b	Overexpression of TtuC(S49V)
pDG71	pET24b	Overexpression of TtuC(A50R)
pDG72	pET24b	Overexpression of TtuC(A56K)
pDG73	pET24b	Overexpression of TtuC(S59G)
pDG74	pET24b	Overexpression of TtuC(E60V)
pDG75	pET24b	Overexpression of TtuC(R72A)
pDG77	pET24b	Overexpression of Rv0100
pDG78	pET24b	Overexpression of MmaA
pRv0099	pET24b	Overexpression of Rv0099
pBamb_5919_KS^0	pET151/D-TOPO	Overexpression of KS^0 of Bamb_5919

Construction of plasmids for protein expression in *E. coli*. Individual genes were PCR amplified from previous lab members' plasmids. All primers are listed in **Table 5-1**. Gibson assembly was used to ligate PCR products to linear pET24b vector. Plasmids were isolated using a Zyppy Miniprep Kit (Zymo Research) and confirmed by DNA sequencing (UC Berkeley DNA Sequencing Facility). The plasmid pBamb_5919_KS^0 was obtained from Greg Challis lab.

Overexpression and purification of recombinant protein. Expression strains were grown in 1 L of LB supplemented with 50 μg/mL of kanamycin at 37 °C and 220 rpm until an OD_{600} of 0.6. Cultures were then cooled on ice for 10 minutes before induction with 200 μM IPTG. Induction of gene expression lasted for 16 h at 16 °C and 200 rpm. The cells were then harvested by centrifugation (6000 x g, 15 min, 4 °C), and supernatant was removed. The pellet was resuspended in 30 mL of lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM imidazole) and lysed with a Q700 sonicator. The insoluble fraction was removed by centrifugation (15,000 rpm, 1 h, 4 °C), and the supernatant was filtered with a 0.45 μM filter before batch binding. Ni-NTA resin (Qiagen) was added to the filtrate at 2 mL/L of cell culture, and samples were allowed to nutate for 1 hour at 4 °C. The protein resin mixture was added to a gravity filter column. The flow through was discarded, the column was then washed with 20-30 mL of wash buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 20 mM imidazole) and tagged protein was eluted in approximately 8 mL of elution buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 250 mM imidazole). Complete elution was determined by Bradford assay. Purified proteins were then concentrated and exchanged into appropriate buffer (50 mM HEPES, pH 7.5, 100 mM NaCl) using Amicon ultra filter units. After two rounds of exchange and concentration, pure protein was removed, and glycerol was added to a final concentration of 8%. Proteins were flash frozen in liquid nitrogen and stored at −80 °C or used immediately for in vitro assays. Presence and purity of enzymes was assessed using SDS-page and concentration was determined using a NanoDrop UV-Vis spectrophotometer (Thermo Scientific).

Heterologous expression of *ttuABC* and *epn*/*epx* pathway in *E. coli*. Two *E. coli* BAP1 strains, one containing pXZ255, pXZ256, pXZ258, and the other containing pXZ255, pXZ256, pXZ260, were grown in 0.5 mL LB medium for seeding overnight at 37 °C with 100 μg/ml carbenicillin, 50 μg/ml spectinomycin, and 50 μg/ml kanamycin. The seed cultures were inoculated into 50 mL LB medium to grow until OD_{600} reached 0.6. Cultures were then cooled on ice for 10 minutes before induction with 500 μM IPTG, fed with 1 mM serine, 1 mM leucine and 1 mM 9-decenoic acid. After 48 h of growth at 18 °C (optimal condition tested), the culture broth and the cell pellets were separated by centrifugation. Compounds were extracted from cell-free supernatant (50ml) and cell pellets using ethyl acetate. The solvent was removed by rotary evaporation. The residue was redissolved in 0.5mL methanol and analyzed by LC/HRMS (20 μL injection). LC/HRMS analysis was performed on an Agilent Technologies 6545 Accurate Mass QTOF LC/MS with an Agilent Eclipse Plus C18 column (4.6 × 100 mm). A linear gradient of 2–98% CH_3CN (v/v) over 30 min in H_2O with 0.1% (v/v) formic acid at a flow rate of 0.5 ml/min was used. LC-HRMS/MS analysis was performed using the same program described above with untargeted mass/mass and a collision energy of 15 eV.

Enzymatic assays in MmaA system. Coupled enzymatic reactions were performed at room temperature for four hours in 50 μL of 50 mM HEPES (pH 8) containing 2.5 mM $MgCl_2$, 5 mM ATP, 1 mM TCEP, 5 mM decanoic acid or 9-decenoic acid, 2 mM lysine, 2 mM NADPH, 20 μM Rv0099/TtuA, 50 μM Rv0100/TtuC, 50 μM MmaA, with and without 50 μM KS^0. Negative controls were conducted by omitting corresponding enzymes. Reactions were quenched with two volumes of cold MeOH, precipitated protein was removed by centrifugation (21,000 x g, 10 min) and the supernatant was used for analysis. LC/HRMS analysis was performed on an Agilent technologies 6545 Accurate-Mass Q-TOF LC-MS instrument with an Agilent Eclipse Plus C18

column (4.6 x 100 mm). A linear gradient of 2-98% CH_3CN with 0.1% formic acid (v/v) over 30 minutes in H_2O with 0.1% formic acid (v/v) with a flow rate of 0.5 mL/min was used.

Chapter 6. Conclusion

Natural products play a critical role in various fields, including pharmaceuticals, agriculture, and materials science, due to their diverse chemical structures and biological activities. Among them, terminal alkene containing natural products and terminal alkyne containing natural products have garnered particular attention due to their unique chemical properties and versatile applications. On one hand, both of them serve as valuable substrates for synthetic modification and potential pharmaceuticals due to their functional group reactivity. The known terminal alkene or alkyne containing PKs/NRPSs exhibit a wide range of bioactivities, including but not limited to anticancer, antimicrobial, and immunosuppressive properties. On the other hand, terminal alkyne containing natural products offer unique chemical functionalities that make them promising candidates for drug discovery and useful tools in chemical biology. The applications of terminal alkynes in bioorthogonal chemistry allow selective labeling, visualization, and quantification of biomolecules in complex biological systems. Hence, the discovery and biosynthesis studies of terminal alkene and alkyne containing natural products hold immense significance. The work presented in the dissertation includes the elucidation and engineering of the biosynthesis of a terminal alkene containing PK/NRP hybrid, the biochemical studies of integral membrane proteins in terminal alkyne formation, and biosynthetic engineering for the generation of terminal alkyne tagged lipopeptides.

In terms of the biosynthesis of terminal alkene containing natural products, we focus on the biosynthesis of salivabactin, which is a recently discovered PK/NRP hybrid with potent antimicrobial activities isolated from the oral probiotic, SAL. Previous investigations into the interactions between SAL and GAS highlighted the potential role of salivabactin in shaping the oral microbiome,[90] suggesting a novel molecular mechanism underlying its bioactivity. Given the numerous oral health benefits associated with this compound, we are interested in its mode of action in antimicrobial activities and intend to generate salivabactin analogs for bioactivity assessment. In the orchestrated assembly line process, the A domain serves as the initial point of interaction for the substrate, preceding its incorporation into the emerging peptide natural product. Adenylation domains facilitate a two-step reaction, activation of the amino acyl substrate into an adenylate, followed by transfer of the amino acid to the thiol of the PPant arm of the carrier protein.[165] Since aromatic acid-activating A domains are often promiscuous, we plan to test the substrate specificity of the A domain in SarD to define the tolerable substrate range, enabling precursor-directed biosynthesis to generate more salivabactin analogs. In the initial feeding trials within the heterologous expression system, the A domain did not accept substrates diverging significantly from 4HBA. Hence, future efforts will focus on engineering the A domain for relaxed substrate specificity. The first two structures of NRPS A domains were PheA, a phenylalanine activating adenylation domain dissected from the multi-domain gramicidin synthetase, and the free-standing 2,3-dihydroxybenzoic acid specific DhbE.[166,167] Both structures indicated that the bulk of the active site is located in the N-terminal subdomain while the Lys on the A10 loop of the C-terminal subdomain is important for the interaction with both the carboxylic acid and the phosphate of the AMP. Therefore, to produce salivabactin analogs, utilizing structural biology insights of the A domain, we can construct docking models and employ mutagenesis to broaden the substrate scope. Additionally, as the BGC of salivabactin distinguishes it among publicly available bacterial genome databases, there are many biosynthetic mechanisms to be elucidated beyond that involved in terminal alkene formation. For instance, we will investigate the timing of

the function of discrete KR and DH domains. We are also interested in the mechanism of the five-membered thiazolidine-ene ring formation and the molecular understanding of the Cy domain in SarE. Furthermore, with the establishment of the heterologous expression in *E. coli*, we aim to enhance the compound titer further by metabolic engineering, such as testing various combinations of replication origins and promoters in plasmids, overexpressing the biosynthetic genes for 4HBA in *E. coli*, and employing multidrug exporters to improve the cell viability.

Apart from the investigations on terminal alkene containing natural products, we also studied the biosynthesis of terminal alkyne formation and the application of the terminal alkyne biosynthetic machinery in *de novo* biosynthesis of alkyne tagged lipopeptides. Only a limited number of enzymes responsible for terminal alkyne formation have been characterized so far. Among them, JamB and its homolog TtuB stand out for several reasons. Firstly, they are the ACP-dependent membrane-bound terminal acetylenases. Secondly, they harbor intriguing mechanistic mysteries in terminal alkyne formation, such as the status of the diiron center, and the bifunctional nature of desaturase and acetylenase. However, the inherent membrane-bound properties of JamB/TtuB significantly impeded the study. We reported our efforts in solubilizing JamB and TtuB with various strategies. On one hand, all solubilized JamB/TtuB showed activity *in vivo*; but to make sure that the activity resulted from the soluble part, it is necessary to conduct activity tests with purified soluble variants of JamB/TtuB. However, only the QTY engineered JamB showed diminished activity *in vitro*, suggesting the importance and urgency of establishing a reliable *in vitro* activity reporting system. The difficulty in the enzymatic assays include the generation of ACP-bound substrates and unknown electron transport partners. On the other, the promise of obtaining soluble JamB/TtuB paves the way for all the subsequent biochemical and structural studies. With purified enzymes, we aim to reconstitute the terminal alkyne formation *in vitro* and establish a platform for in-depth characterization of desaturation/acetylenation process, including scrutinizing the electron transport chain, studying the enzyme kinetics, and capturing potential intermediates. The diiron center status during terminal alkyne formation will be explored by spectroscopic methods, such as electron paramagnetic resonance (EPR), X-ray absorbance spectroscopy (XAS) and Mössbauer spectroscopy. In addition, mechanism elucidation will be coupled with structural biology investigations, including X-ray crystallography and single-particle cryo-electron microscopy (cryo-EM), which reveals active sites, substrate binding tunnels and electron partner interfaces. Meanwhile, with the rapid development of artificial intelligence in structural biology, we can refine existing models and examine the structure-function relationship with *in vitro* enzymatic assays. Furthermore, based on the understanding of terminal alkyne biosynthesis, we aim to tag biomolecules with the terminal alkyne functionality through biosynthetic engineering. The generation of terminal alkyne tagged lipopeptides remained relatively underexplored, primarily due to the limited understanding of the interactions between ACPs and NRPSs. Interactions between carrier proteins, including ACPs and PCPs, and other biosynthetic enzyme domains are transient in nature, posing challenges in studying their specific interfaces in bond formation. Despite predictions derived from structural modeling and docking algorithms, the anticipated enhancement of TtuC recognition by MmaA through rational design of mutants did not materialize as expected. This emphasizes the complexity of carrier protein regulation in natural systems. Our next step involves random mutagenesis to delve deeper into the structure-function relationship, ultimately aiming to achieve enhanced recognition between JamC/TtuC and NRPSs. Moreover, we plan to use hydrogen/deuterium exchange mass spectrometry (HDX-MS) analysis, which has proven to be powerful in mapping dynamic protein binding interactions, to probe the recognition between upstream ACPs and downstream NRPSs.

In conclusion, the dissertation encompasses comprehensive investigations into the understanding and engineering of the biosynthesis of terminal alkene and alkyne containing natural products. The studies on terminal alkene containing natural products focus on a new antibiotic, salivabactin. We elucidated the biosynthetic pathway, characterized unique biosynthetic enzymes and subsequently engineered the pathway to enhance the production of salivabactin. For terminal alkyne containing natural products, we delved into the biochemical studies focusing on the family of ACP-dependent membrane-bound desaturases/acetylenases, facilitating the understanding of mechanistic intricacies of the terminal alkyne formation. Furthermore, we explored biosynthetic engineering with the goal of generating terminal alkyne tagged lipopeptides, offering innovative methodologies to manipulate natural product biosynthesis. The work presented in the dissertation not only advances our fundamental understandings of terminal alkene and alkyne biosynthesis, but also expands the applications of biosynthetic engineering, fostering the development of novel therapeutics and biotechnological tools.

References

1. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
2. Dias, D. A., Urban, S. & Roessner, U. A Historical overview of natural products in drug discovery. *Metabolites* **2**, 303–336 (2012).
3. Thomford, N. E. *et al.* Natural products for drug discovery in the 21st century: Innovations for novel drug discovery. *Int. J. Mol. Sci.* **19**, (2018).
4. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
5. Platon, V. M., Dragoi, B. & Marin, L. Erythromycin Formulations—A Journey to Advanced Drug Delivery. *Pharmaceutics* **14**, 2180 (2022).
6. Law, B. K. Rapamycin: An anti-cancer immunosuppressant? *Crit. Rev. Oncol. Hematol.* **56**, 47–60 (2005).
7. Weaver, B. A. How Taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **25**, 2677–2681 (2014).
8. J, A. *et al.* Quinine, an old anti-malarial drug in a modern world: Role in the treatment of malaria. *Malar. J.* **10**, 144 (2011).
9. Dayan, F. E., Cantrell, C. L. & Duke, S. O. Natural products in crop protection. *Bioorg. Med. Chem.* **17**, 4022–4034 (2009).
10. González-Manzano, S. & Dueñas, M. Applications of natural products in food. *Foods* **10**, 300 (2021).
11. Chen, Y. *et al.* Dietary Supplements and Natural Products: An Update on Their Clinical Effectiveness and Molecular Mechanisms of Action During Accelerated Biological Aging. *Front. Genet.* **13**, 880421 (2022).
12. Liu, J. K. Natural products in cosmetics. *Nat. Products Bioprospect.* **12**, 40 (2022).
13. Ige, O. O., Umoru, L. E. & Aribo, S. Natural Products: A Minefield of Biomaterials. *ISRN Mater. Sci.* **2012**, 983062 (2012).
14. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).
15. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127–135 (2016).
16. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
17. Ninkuu, V. *et al.* Biochemistry of Terpenes and Recent Advances in Plant Protection. *International Journal of Molecular Sciences* vol. 22 (2021).
18. Elshahawi, S. I., Shaaban, K. A., Kharel, M. K. & Thorson, J. S. A comprehensive review of glycosylated bacterial natural products. *Chem. Soc. Rev.* **44**, 7591–7697 (2015).
19. Kries, H. & O'Connor, S. E. Biocatalysts from alkaloid producing plants. *Curr. Opin. Chem. Biol.* **31**, 22–30 (2016).
20. Fisch, K. M. Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS. *RSC Adv.* **3**, 18228–18247 (2013).
21. Sattely, E. S., Fischbach, M. A. & Walsh, C. T. Total biosynthesis: In vitro reconstitution of polyketide and nonribosomal peptide pathways. *Nat. Prod. Rep.* **25**, 757–793 (2008).

22. Liu, J., Zhu, X., Kim, S. J. & Zhang, W. Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities. *Nat. Prod. Rep.* **33**, 1146–1165 (2016).
23. Sundaram, S. & Hertweck, C. On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems. *Curr. Opin. Chem. Biol.* **31**, 82–94 (2016).
24. Porterfield, W. B. & Zhang, W. *Mutasynthesis for Natural Product Bioengineering*. *Comprehensive Natural Products III* (Elsevier Ltd., 2020). doi:10.1016/b978-0-12-409547-2.14633-5.
25. Keatinge-Clay, A. T. The structures of type i polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
26. Gulick, A. M. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827 (2009).
27. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chemie Int. Ed.* **48**, 4688–4716 (2009).
28. Beck, C., Garzón, J. F. G. & Weber, T. Recent Advances in Re-engineering Modular PKS and NRPS Assembly Lines. *Biotechnol. Bioprocess Eng.* **25**, 886–894 (2020).
29. Moschny, J. *et al.* Precursor-Directed Biosynthesis and Fluorescence Labeling of Clickable Microcystins. *J. Nat. Prod.* **83**, 1960–1970 (2020).
30. Huang, D., Xia, M., Li, S., Wen, J. & Jia, X. Enhancement of FK506 production by engineering secondary pathways of Streptomyces tsukubaensis and exogenous feeding strategies. *J. Ind. Microbiol. Biotechnol.* **40**, 1023–1037 (2013).
31. Andexer, J. N. *et al.* Biosynthesis of the immunosuppressants FK506, FK520, and rapamycin involves a previously undescribed family of enzymes acting on chorismate. *Proc. Natl. Acad. Sci.* **108**, 4776–4781 (2011).
32. Goss, R. J. M. *et al.* An Expeditious Route to Fluorinated Rapamycin Analogues by Utilising Mutasynthesis. *ChemBioChem* **11**, 698–702 (2010).
33. Ban, Y. H. *et al.* Mutational biosynthesis of a FK506 analogue containing a non-natural starter unit. *Mol. Biosyst.* **9**, 944–947 (2013).
34. Mo, S. *et al.* Biosynthesis of the Allylmalonyl-CoA Extender Unit for the FK506 Polyketide Synthase Proceeds through a Dedicated Polyketide Synthase and Facilitates the Mutasynthesis of Analogues. *J. Am. Chem. Soc.* **133**, 976–985 (2011).
35. Lechner, A. *et al.* Designed Biosynthesis of 36-Methyl-FK506 by Polyketide Precursor Pathway Engineering. *ACS Synth. Biol.* **2**, 379–383 (2013).
36. Awakawa, T. *et al.* Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution. *Nat. Commun.* **9**, 3534 (2018).
37. Watanabe, K. *et al.* Rationally Engineered Total Biosynthesis of a Synthetic Analogue of a Natural Quinomycin Depsipeptide in Escherichia coli. *ChemBioChem* **10**, 1965–1968 (2009).
38. Koch, A. A. *et al.* A Single Active Site Mutation in the Pikromycin Thioesterase Generates a More Effective Macrocyclization Catalyst. *J. Am. Chem. Soc.* **139**, 13456–13465 (2017).
39. Hopwood, D. A. *et al.* Production of 'hybrid' antibiotics by genetic engineering. *Nature* **314**, 642–644 (1985).
40. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
41. Klaus, M. *et al.* Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *J. Biol. Chem.* **291**, 16404–

16415 (2016).
42. J., C. M., G., O. J., L., L. I. & F., A. D. Biosynthesis of Novel Pyoverdines by Domain Substitution in a Nonribosomal Peptide Synthetase of Pseudomonas aeruginosa . *Appl. Environ. Microbiol.* **80**, 5723–5731 (2014).
43. Wang, Z. *et al.* De Novo Design and Implementation of a Tandem Acyl Carrier Protein Domain in a Type I Modular Polyketide Synthase. *ACS Chem. Biol.* **13**, 3072–3077 (2018).
44. Yin, X., Chen, Y., Zhang, L., Wang, Y. & Zabriskie, T. M. Enduracidin Analogues with Altered Halogenation Patterns Produced by Genetically Engineered Strains of Streptomyces fungicidicus. *J. Nat. Prod.* **73**, 583–589 (2010).
45. Jensen, C. M. & Trogler, W. C. Catalytic Hydration of Terminal Alkenes to Primary Alcohols. *Science* **233**, 1069–1071 (1986).
46. Copéret, C., Adolfsson, H. & Barry Sharpless, K. A simple and efficient method for epoxidation of terminal alkenes. *Chem. Commun.* 1565–1566 (1997) doi:10.1039/A703542J.
47. Nishimura, T., Kakiuchi, N., Onoue, T., Ohe, K. & Uemura, S. Palladium(II)-catalyzed oxidation of terminal alkenes to methyl ketones using molecular oxygen. *J. Chem. Soc. Perkin Trans. 1* 1915–1918 (2000) doi:10.1039/B001854F.
48. Mlynarski, S. N., Schuster, C. H. & Morken, J. P. Asymmetric synthesis from terminal alkenes by cascades of diboration and cross-coupling. *Nature* **505**, 386–390 (2014).
49. Klar, U. *et al.* Total Synthesis and Antitumor Activity of ZK-EPO: The First Fully Synthetic Epothilone in Clinical Development. *Angew. Chemie Int. Ed.* **45**, 7942–7948 (2006).
50. KINO, T. *et al.* FK-506, a novel immunosuppressant isolated from a Streptomyces. I. Fermentation, isolation, and physico-chemical and biological characteristics. *J. Antibiot. (Tokyo).* **40**, 1249–1255 (1987).
51. Fudou, R., Iizuka, T. & Yamanaka, S. Haliangicin, a novel antifungal metabolite produced by a marine myxobacterium. 1. Fermentation and biological characteristics. *J. Antibiot. (Tokyo).* **54**, 149–152 (2001).
52. Chang, Z. *et al.* Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **67**, 1356–1367 (2004).
53. Mitsuhashi, S. *et al.* Tautomycetin Is a Novel and Specific Inhibitor of Serine/Threonine Protein Phosphatase Type 1, PP1. *Biochem. Biophys. Res. Commun.* **287**, 328–331 (2001).
54. Rui, Z. *et al.* Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18237–18242 (2014).
55. Gu, L. *et al.* Polyketide decarboxylative chain termination preceded by O-sulfonation in curacin A biosynthesis. *J. Am. Chem. Soc.* **131**, 16033–16035 (2009).
56. Gehret, J. J. *et al.* Terminal alkene formation by the thioesterase of curacin A biosynthesis: Structure of a decarboxylating thioesterase. *J. Biol. Chem.* **286**, 14445–14454 (2011).
57. Li, W. *et al.* Characterization of the Tautomycetin Biosynthetic Gene Cluster from Streptomyces griseochromogenes Provides New Insight into Dialkylmaleic Anhydride Biosynthesis. *J. Nat. Prod.* **72**, 450–459 (2009).
58. Scaglione, J. B. *et al.* Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. *Angew. Chemie Int. Ed.* **49**, 5726–5730 (2010).
59. Sun, Y. *et al.* Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci.*

Rep. **6**, 22091 (2016).
60. Blake-Hedges, J. M. *et al.* Structural Mechanism of Regioselectivity in an Unusual Bacterial Acyl-CoA Dehydrogenase. *J. Am. Chem. Soc.* **142**, 835–846 (2020).
61. Dennig, A. *et al.* Oxidative Decarboxylation of Short-Chain Fatty Acids to 1-Alkenes. *Angew. Chemie Int. Ed.* **54**, 8819–8822 (2015).
62. Rui, Z., Harris, N. C., Zhu, X., Huang, W. & Zhang, W. Discovery of a Family of Desaturase-Like Enzymes for 1-Alkene Biosynthesis. *ACS Catal.* **5**, 7091–7094 (2015).
63. Huang, Q., Wang, W.-N. & Zhu, S.-F. Iron-Catalyzed Alkylzincation of Terminal Alkynes. *ACS Catal.* **12**, 2581–2588 (2022).
64. Binder, W. & Kluger, C. Azide/Alkyne-“Click” Reactions: Applications in Material Science and Organic Synthesis. *Curr. Org. Chem.* **10**, 1791–1815 (2006).
65. Hein, C. D., Liu, X.-M. & Wang, D. Click Chemistry, A Powerful Tool for Pharmaceutical Sciences. *Pharm. Res.* **25**, 2216–2230 (2008).
66. Scinto, S. L. *et al.* Bioorthogonal chemistry. *Nat. Rev. Methods Prim.* **1**, 30 (2021).
67. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie Int. Ed.* **40**, 2004–2021 (2001).
68. Gu, D. & Zhang, W. Engineered biosynthesis of alkyne-tagged polyketides. *Methods Enzymol.* **665**, 347–373 (2022).
69. Mullins, A. J. *et al.* Discovery of the Pseudomonas Polyyne Protegencin by a Phylogeny-Guided Study of Polyyne Biosynthetic Gene Cluster Diversity. *MBio* **12**, e00715-21 (2021).
70. Li, X., Lv, J. M., Hu, D. & Abe, I. Biosynthesis of alkyne-containing natural products. *RSC Chem. Biol.* **2**, 166–180 (2021).
71. Rio, D. *et al.* Biosynthesis of Isonitrile- and Alkyne-Containing Natural Products. *Annu. Rev. Chem. Biomol. Eng* **13**, 1–24 (2022).
72. Chai, Q.-Y., Yang, Z., Lin, H.-W. & Han, B.-N. Alkynyl-Containing Peptides of Marine Origin: A Review. *Marine Drugs* vol. 14 216 (2016).
73. Zhu, X., Liu, J. & Zhang, W. De novo biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11**, 115–120 (2015).
74. Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019).
75. Shanklin, J., Guy, J. E. & Mishra, G. Desaturases : Emerging Models for Understanding Functional Diversification of Diiron-containing Enzymes. *J. Biol. Chem.* **284**, 18559–18563 (2009).
76. Shanklin, J. & Cahoon, E. B. Desaturation and related modifications of fatty acids. *Annu. Rev. Plant Biol.* **49**, 611–641 (1998).
77. Zhu, X., Su, M., Manickam, K. & Zhang, W. Bacterial Genome Mining of Enzymatic Tools for Alkyne Biosynthesis. *ACS Chem. Biol.* **10**, 2785–2793 (2015).
78. Porterfield, W. B., Poenateetai, N. & Zhang, W. Engineered Biosynthesis of Alkyne-Tagged Polyketides by Type I PKSs. *iScience* **23**, 100938 (2020).
79. Dekaboruah, E., Suryavanshi, M. V., Chettri, D. & Verma, A. K. Human microbiome: an academic update on human body site specific surveillance and its possible role. *Arch. Microbiol.* **202**, 2147–2167 (2020).
80. J.A., L. *et al.* The Biology of Streptococcus mutans. *Microbiol. Spectr.* **7**, 10.1128/microbiolspec.gpp3-0051–2018 (2019).
81. Carapetis, J. R., Steer, A. C., Mulholland, E. K. & Weber, M. The global burden of group A streptococcal diseases. *Lancet Infect. Dis.* **5**, 685–694 (2005).

82. Brouwer, S. *et al.* Pathogenesis, epidemiology and control of Group A Streptococcus infection. *Nat. Rev. Microbiol.* **21**, 431–447 (2023).
83. Caroline, B., Pablo, A.-M., Francis, F., Pierre, R. & Bernard, B. Genome Sequence of the Lantibiotic Bacteriocin Producer Streptococcus salivarius Strain K12. *J. Bacteriol.* **194**, 5959–5960 (2012).
84. Simone, G. *et al.* Oral Bacteria as Potential Probiotics for the Pharyngeal Mucosa. *Appl. Environ. Microbiol.* **76**, 3948–3958 (2010).
85. He, L., Yang, H., Chen, Z. & Ouyang, X. The Effect of Streptococcus salivarius K12 on Halitosis: a Double-Blind, Randomized, Placebo-Controlled Trial. *Probiotics Antimicrob. Proteins* **12**, 1321–1329 (2020).
86. Barber, C. C. & Zhang, W. Small molecule natural products in human nasal/oral microbiota. *J. Ind. Microbiol. Biotechnol.* **48**, kuab010 (2021).
87. Hu, Z. & Zhang, W. Signaling Natural Products from Human Pathogenic Bacteria. *ACS Infect. Dis.* **6**, 25–33 (2020).
88. Gajender, A. *et al.* Identification of the Bacterial Biosynthetic Gene Clusters of the Oral Microbiome Illuminates the Unexplored Social Language of Bacteria during Health and Disease. *MBio* **10**, 10.1128/mbio.00321-19 (2019).
89. Zupancic, K., Kriksic, V., Kovacevic, I. & Kovacevic, D. Influence of Oral Probiotic Streptococcus salivarius K12 on Ear and Oral Cavity Health in Humans: Systematic Review. *Probiotics Antimicrob. Proteins* **9**, 102–110 (2017).
90. Do, H. *et al.* Engineered probiotic overcomes pathogen defences using signal interference and antibiotic production to treat infection in mice. *Nat. Microbiol.* **9**, 502–513 (2024).
91. Otto, H. *et al.* Salivaricin A2 and the Novel Lantibiotic Salivaricin B Are Encoded at Adjacent Loci on a 190-Kilobase Transmissible Megaplasmid in the Oral Probiotic Strain Streptococcus salivarius K12. *Appl. Environ. Microbiol.* **73**, 1107–1113 (2007).
92. Bloudoffa, K., Fageb, C. D., Marahiel, M. A. & Schmeinga, T. M. Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 95–100 (2017).
93. Zhang, F. *et al.* MS-Derived Isotopic Fine Structure Reveals Forazoline A as a Thioketone-Containing Marine-Derived Natural Product. *Org. Lett.* **22**, 1275–1279 (2020).
94. Passmore, M., Gallo, A., Lewandowski, J. R. & Jenner, M. Molecular basis for acyl carrier protein-ketoreductase interaction intrans-acyltransferase polyketide synthases. *Chem. Sci.* **12**, 13676–13685 (2021).
95. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **98**, 7735–7746 (2014).
96. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: Catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31**, 61–108 (2014).
97. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **291**, 1790–1792 (2001).
98. Tsai, S. C. *et al.* 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**, 14808–14813 (2001).
99. Caswell, B. T. *et al.* Thioesterase enzyme families: Functions, structures, and mechanisms.

Protein Sci. **31**, 652–676 (2022).
100. Wei, Q., Wang, Z.-P., Zhang, X. & Zou, Y. Diaryl Ether Formation by a Versatile Thioesterase Domain. *J. Am. Chem. Soc.* **144**, 9554–9558 (2022).
101. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, e4519 (2023).
102. Galea, C. A. *et al.* Functional characterization of the unique terminal thioesterase domain from polymyxin synthetase. *Biochemistry* **56**, 657–668 (2017).
103. Swarbrick, C. M. D., Nanson, J. D., Patterson, E. I. & Forwood, J. K. Structure, function, and regulation of thioesterases. *Prog. Lipid Res.* **79**, 101036 (2020).
104. Kreitler, D. F., Gemmell, E. M., Schaffer, J. E., Wencewicz, T. A. & Gulick, A. M. The structural basis of N-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nature Communications* vol. 10 3432 (2019).
105. Oliver, R. A., Li, R. & Townsend, C. A. Monobactam formation in sulfazecin by a nonribosomal peptide synthetase thioesterase. *Nat. Chem. Biol.* **14**, 5–7 (2018).
106. Mir Mohseni, M. *et al.* Discovery of a Mosaic-Like Biosynthetic Assembly Line with a Decarboxylative Off-Loading Mechanism through a Combination of Genome Mining and Imaging. *Angew. Chemie Int. Ed.* **55**, 13611–13614 (2016).
107. Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12432–12437 (2016).
108. Wyche, T. P. *et al.* Forazoline A: Marine-Derived Polyketide with Antifungal In Vivo Efficacy. *Angew. Chemie Int. Ed.* **53**, 11583–11586 (2014).
109. Rahman, A. S., Hothersall, J., Crosby, J., Simpson, T. J. & Thomas, C. M. Tandemly Duplicated Acyl Carrier Proteins, Which Increase Polyketide Antibiotic Production, Can Apparently Function Either in Parallel or in Series*. *J. Biol. Chem.* **280**, 6399–6408 (2005).
110. Thongkawphueak, T. *et al.* Solution Structure and Conformational Dynamics of a Doublet Acyl Carrier Protein from Prodigiosin Biosynthesis. *Biochemistry* **60**, 219–230 (2021).
111. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412–1436 (2019).
112. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **252**, 675–679 (1991).
113. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges . *Frontiers in Microbiology* vol. 5 172 (2014).
114. Pfeifer, B., Hu, Z., Licari, P. & Khosla, C. Process and Metabolic Strategies for Improved Production of Escherichia coli-Derived 6-Deoxyerythronolide B. *Appl. Environ. Microbiol.* **68**, 3287–3292 (2002).
115. Zhao, S. *et al.* Heterologous expression facilitates the discovery and characterization of marine microbial natural products. *Eng. Microbiol.* **4**, 100137 (2024).
116. Zhang, N. & An, Z. Heterologous Protein Expression in Yeasts and Filamentous Fungi. in *Manual of Industrial Microbiology and Biotechnology* 145–156 (2010). doi:https://doi.org/10.1128/9781555816827.ch11.
117. Dormiani, K., Khazaie, Y., Sadeghi, H. M. M., Rabbani, M. & Moazen, F. Cloning and expression of a human tissue plasminogen activator variant: K2S in Escherichia coli. *Pakistan J. Biol. Sci. PJBS* **10**, 946–949 (2007).
118. Su, M., Zhu, X. & Zhang, W. Probing the acyl carrier protein–Enzyme interactions within terminal alkyne biosynthetic machinery. *AIChE J.* **64**, 4255–4262 (2018).
119. Zhu, X., Shieh, P., Su, M., Bertozzi, C. R. & Zhang, W. A fluorogenic screening platform

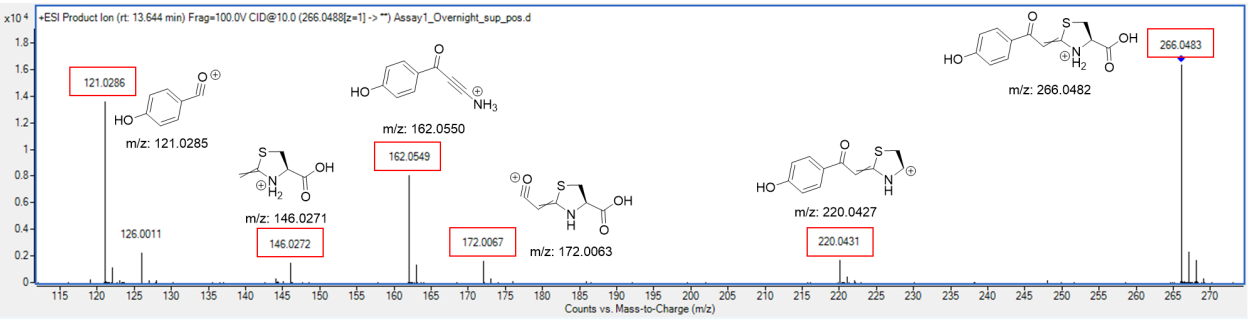
enables directed evolution of an alkyne biosynthetic tool. *Chem. Commun.* **52**, 11239–11242 (2016).
120. Morita, H. *et al.* Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase. *PNAS* **108**, 13504–13509 (2011).
121. Bai, Y. *et al.* X-ray structure of a mammalian stearoyl-CoA desaturase. *Nature* **524**, 252–256 (2015).
122. Wang, H. *et al.* Crystal structure of human stearoyl-coenzyme A desaturase in complex with substrate. *Nat. Struct. Mol. Biol.* **22**, 581–585 (2015).
123. Shen, J., Wu, G., Tsai, A. L. & Zhou, M. Structure and Mechanism of a Unique Diiron Center in Mammalian Stearoyl-CoA Desaturase. *J. Mol. Biol.* **432**, 5152–5161 (2020).
124. Ratkeviciute, G., Cooper, B. F. & Knowles, T. J. Methods for the solubilisation of membrane proteins: The micelle-aneous world of membrane protein solubilisation. *Biochem. Soc. Trans.* **49**, 1763–1777 (2021).
125. Denisov, I. G. & Sligar, S. G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 481–486 (2016).
126. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Lett.* **2**, 853–856 (2002).
127. Denisov, I. G., Grinkova, Y. V, Lazarides, A. A. & Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **126**, 3477–3487 (2004).
128. Zhang, S. *et al.* QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8652–E8659 (2018).
129. Panganiban, B. *et al.* Random heteropolymers preserve protein function in Foreign Environments. *Science* **359**, 1239–1243 (2018).
130. Schoborg, J. A. *et al.* A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnol. Bioeng.* **115**, 739–750 (2018).
131. Mizrachi, D. *et al.* RETRACTED ARTICLE: A water-soluble DsbB variant that catalyzes disulfide-bond formation in vivo. *Nat. Chem. Biol.* **13**, 1022–1028 (2017).
132. Schulte-Zweckel, J. *et al.* A hydroxylamine probe for profiling: S-acylated fatty acids on proteins. *Chem. Commun.* **55**, 11183–11186 (2019).
133. Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: Applications come of age. *Biotechnol. Adv.* **30**, 1185–1194 (2012).
134. Li, J. *et al.* Cell-free protein synthesis enables high yielding synthesis of an active multicopper oxidase. *Biotechnol. J.* **11**, 212–218 (2016).
135. Stark, J. C. *et al.* BioBits^{TM} Bright: A fluorescent synthetic biology education kit. *Sci. Adv.* **4**, eaat5107 (2018).
136. Mizrachi, D. *et al.* Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy. *Nat. Commun.* **6**, 6826 (2015).
137. Iqbal, T., Murugan, S., Rajendran, K., Sidhu, J. S. & Das, D. Unraveling the Conversion of Fatty Acids into Terminal Alkenes by an Integral Membrane Enzyme, UndB. *ACS Catal.* **13**, 15516–15525 (2023).
138. Meluzzi, D., Zheng, W. H., Hensler, M., Nizet, V. & Dorrestein, P. C. Top-down mass spectrometry on low-resolution instruments: Characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorganic Med. Chem. Lett.* **18**, 3107–3111 (2008).
139. Zhang, W., Bolla, M. L., Kahne, D. & Walsh, C. T. A three enzyme pathway for 2-amino-

3-hydroxycyclopent-2-enone formation and incorporation in natural product biosynthesis. *J. Am. Chem. Soc.* **132**, 6402–6411 (2010).
140. Levine, M. Z., Gregorio, N. E., Jewett, M. C., Watts, K. R. & Oza, J. P. Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. *J. Vis. Exp.* **2019**, 1–11 (2019).
141. DeGuire, S. M. *et al.* Fluorescent probes of the apoptolidins and their utility in cellular localization studies. *Angew. Chemie - Int. Ed.* **54**, 961–964 (2015).
142. Seidel, J. *et al.* Structure-activity-distribution relationship study of anti-cancer antimycin-type depsipeptides. *Chem. Commun.* **55**, 9379–9382 (2019).
143. Brewitz, L. *et al.* Alkyne Derivatives of SARS-CoV-2 Main Protease Inhibitors Including Nirmatrelvir Inhibit by Reacting Covalently with the Nucleophilic Cysteine. *J. Med. Chem.* **66**, 2663–2680 (2023).
144. Staub, I. & Sieber, S. A. β-Lactams as Selective Chemical Probes for the in Vivo Labeling of Bacterial Enzymes Involved in Cell Wall Biosynthesis, Antibiotic Resistance, and Virulence. *J. Am. Chem. Soc.* **130**, 13400–13409 (2008).
145. Morita, H. *et al.* Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase. *Proc. Natl. Acad. Sci.* **108**, 13504–13509 (2011).
146. Yu, D., Xu, F., Zeng, J. & Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285–295 (2012).
147. Wlodek, A. *et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Commun.* **8**, 1206 (2017).
148. C., B. *et al.* Biosynthetic Gene Cluster for the Polyenoyltetramic Acid α-Lipomycin. *Antimicrob. Agents Chemother.* **50**, 2113–2121 (2006).
149. Rawlings, B. J. Type I polyketide biosynthesis in bacteria (Part A—erythromycin biosynthesis). *Nat. Prod. Rep.* **18**, 190–227 (2001).
150. Zeng, J. *et al.* Portability and Structure of the Four-Helix Bundle Docking Domains of trans-Acyltransferase Modular Polyketide Synthases. *ACS Chem. Biol.* **11**, 2466–2474 (2016).
151. Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci.* **109**, 4110–4115 (2012).
152. Yan, Y. *et al.* Multiplexing of Combinatorial Chemistry in Antimycin Biosynthesis: Expansion of Molecular Diversity and Utility. *Angew. Chemie Int. Ed.* **52**, 12308–12312 (2013).
153. Khem Raj, M. & Shamsher S, K. Lipopeptides as the antifungal and antibacterial agents: applications in food safety and therapeutics. *Biomed Res. Int.* **2015**, 473050 (2015).
154. Liu, J., Zhu, X. & Zhang, W. Identifying the Minimal Enzymes Required for Biosynthesis of Epoxyketone Proteasome Inhibitors. *ChemBioChem* **16**, 2585–2589 (2015).
155. Harris, N. C. *et al.* Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7025–7030 (2017).
156. Schorn, M. *et al.* Genetic Basis for the Biosynthesis of the Pharmaceutically Important Class of Epoxyketone Proteasome Inhibitors. *ACS Chem. Biol.* **9**, 301–309 (2014).
157. Del Rio Flores, A. *et al.* Biosynthesis of Isonitrile Lipopeptide Metallophores from Pathogenic Mycobacteria. *Biochemistry* **62**, 824–834 (2023).
158. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).
159. Masschelein, J. *et al.* A dual transacylation mechanism for polyketide synthase chain release

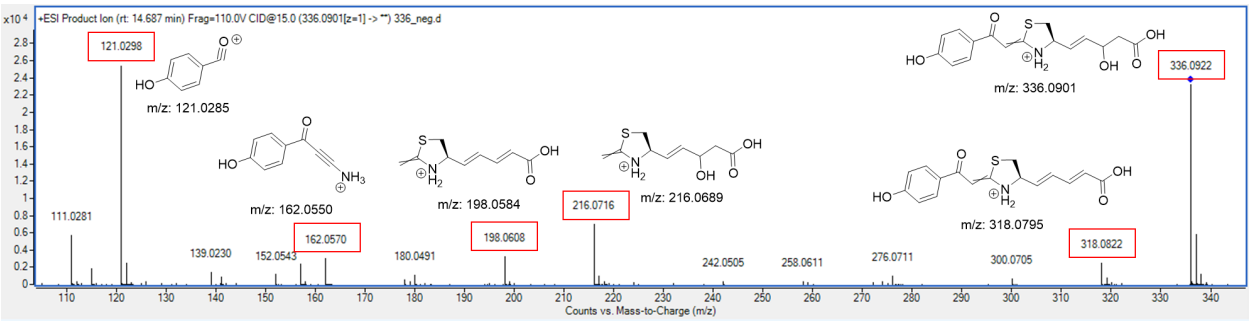
in enacyloxin antibiotic biosynthesis. *Nat. Chem.* **11**, 906–912 (2019).
160. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *J. Mol. Biol.* **425**, 3137–3150 (2013).
161. Bloudoff, K., Alonzo, D. A. & Schmeing, T. M. Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* **23**, 331–339 (2016).
162. Corpuz, J. C., Sanlley, J. O. & Burkart, M. D. Protein-protein interface analysis of the non-ribosomal peptide synthetase peptidyl carrier protein and enzymatic domains. *Synth. Syst. Biotechnol.* **7**, 677–688 (2022).
163. Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663 (2008).
164. Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, (2019).
165. Miller, B. R. & Gulick, A. M. Structural Biology of Nonribosomal Peptide Synthetases BT - Nonribosomal Peptide and Polyketide Biosynthesis: Methods and Protocols. in (ed. Evans, B. S.) 3–29 (Springer New York, 2016). doi:10.1007/978-1-4939-3375-4_1.
166. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
167. May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12120–12125 (2002).

Appendices

A. MS/MS fragmentation patterns of compound **3** and **335**.

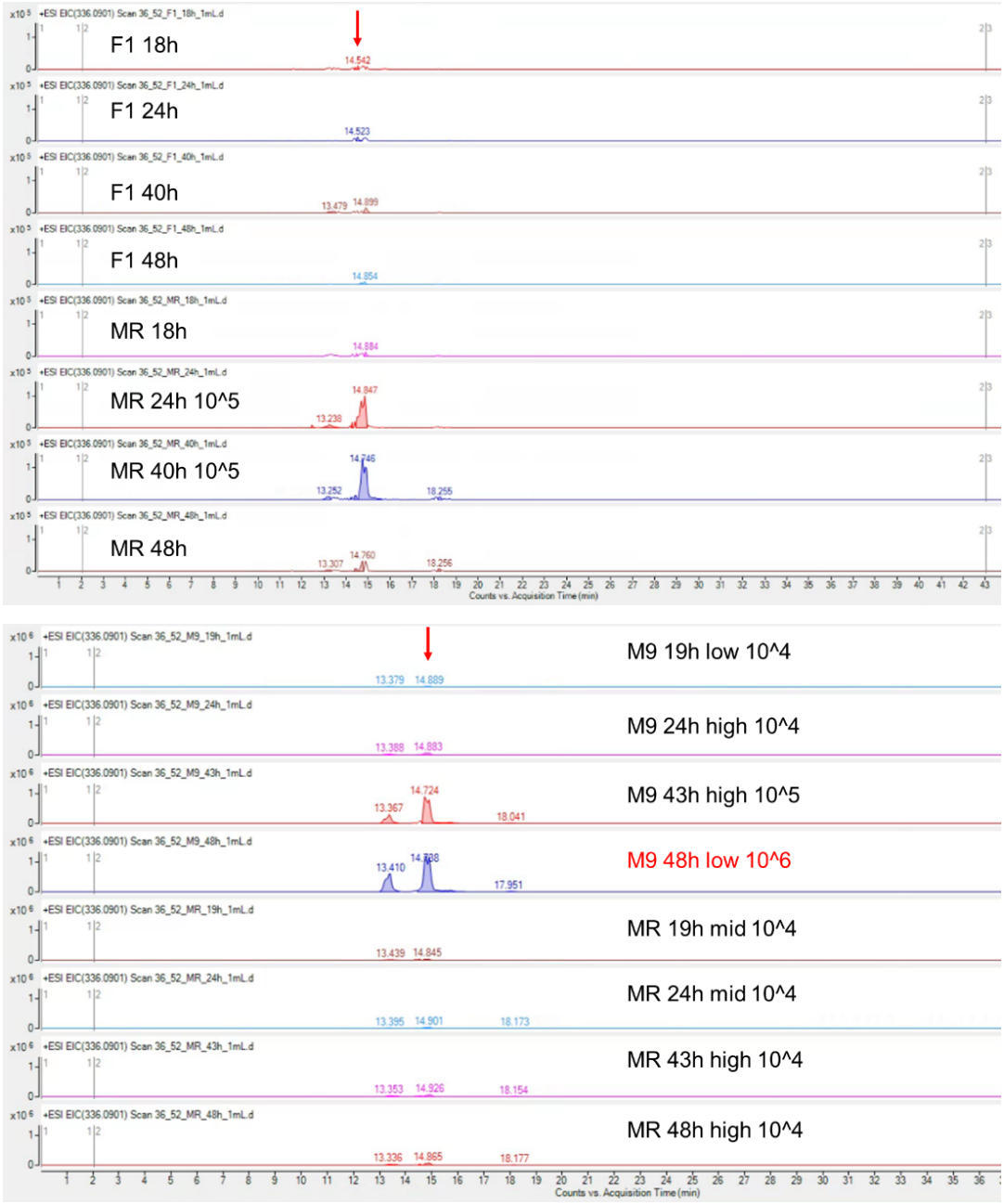


MS/MS fragmentation pattern of compound **3**



MS/MS fragmentation pattern of compound **335**

B. Optimization of compound 335 production in *E. coli*. LC/HRMS measurement with 1 mL culture equivalent injection.

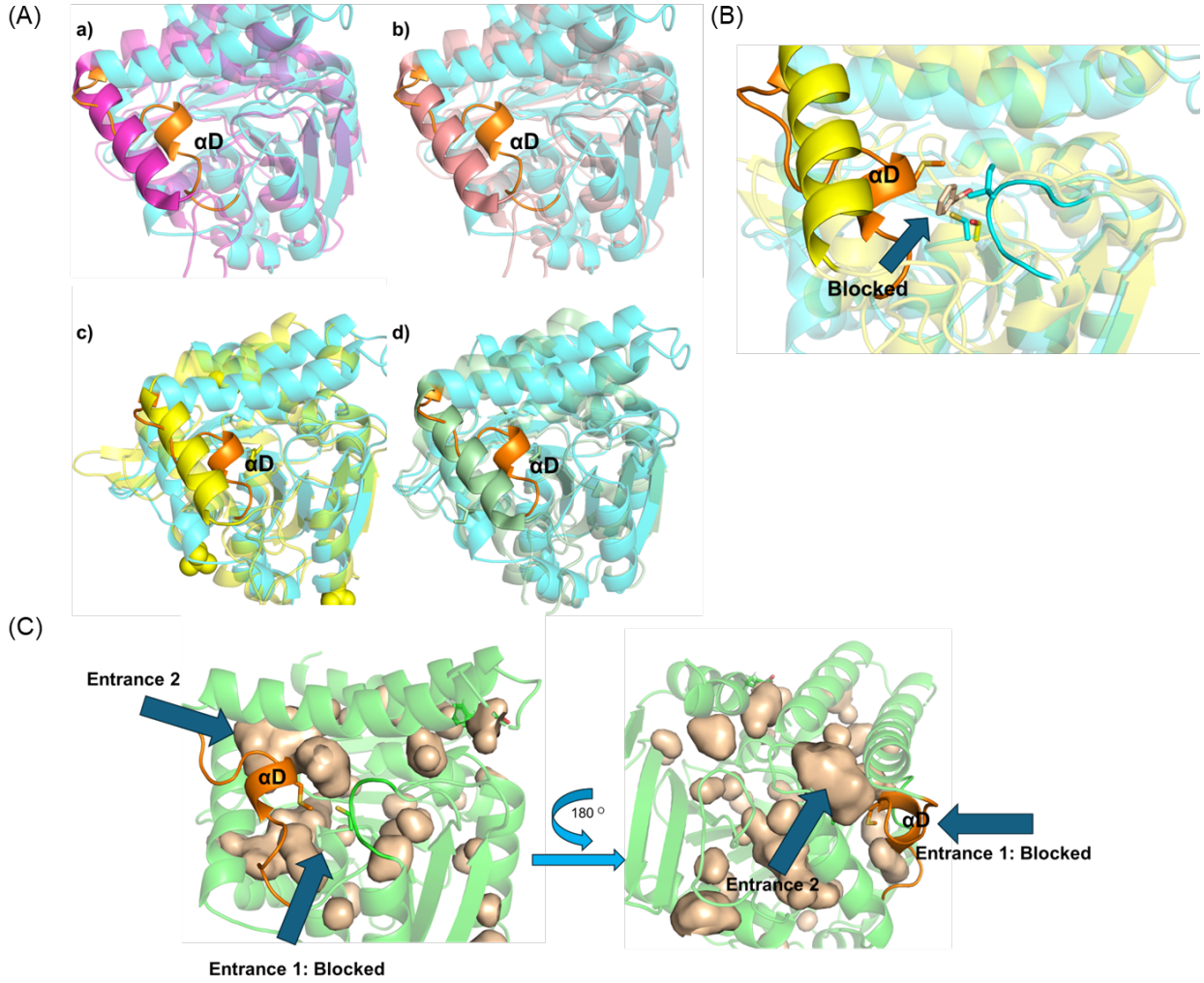


C. Sequence alignment of ACP2 and ACP3 in SarH.

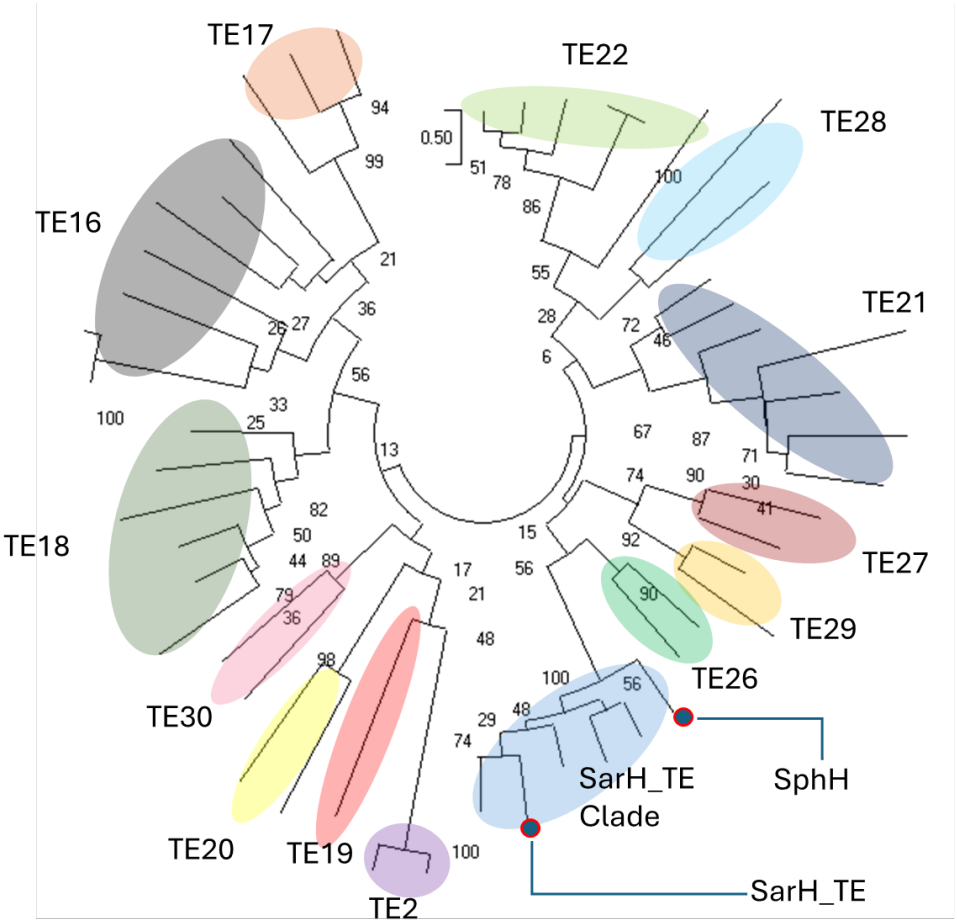
```
ACP2        DDVLKVFSDVLAIPLERLEVDRDIASFGVDSIMLSTLTKEFKRIFPTFEERILIENSTIS      60
ACP3        KKVLRIFSKTLDITSDRLDVDNDIASFGVDSIMLSTLTKEFEKAFPEFEERILVENNTIR      60
            ..**::**..* *  :**:**.********************:: ** ******:**.**

ACP2        E       61
ACP3        E       61
            *
```

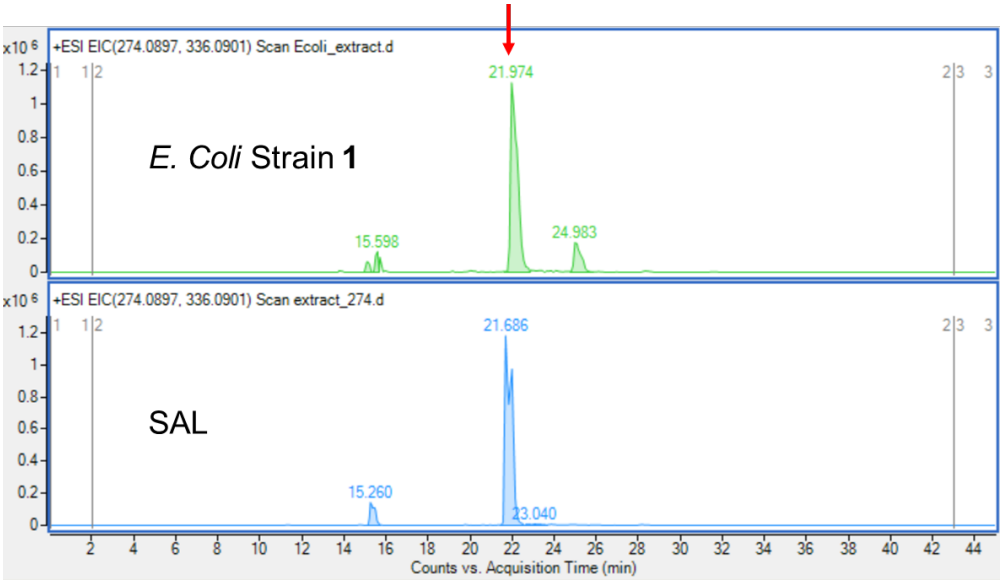
D. Alignment of SarH_TE with Structural homologs given by Dali Server. (A) SarH_TE are colored cyan, with αD highlighted in orange. Overall structure of the structural homologs and SarH_TE aligns well, with an RMSD of a)1.8 Å, b) 1.72 Å, c) 1.88 Å, d) 1.89 Å, respectively. (B) Alignment with 1A8U shows that αD in SarH_TE blocks the substrate access to catalytic center. (C) Substrates would access the catalytic center via entrance **2** since entrance **1** is blocked by αD motif.



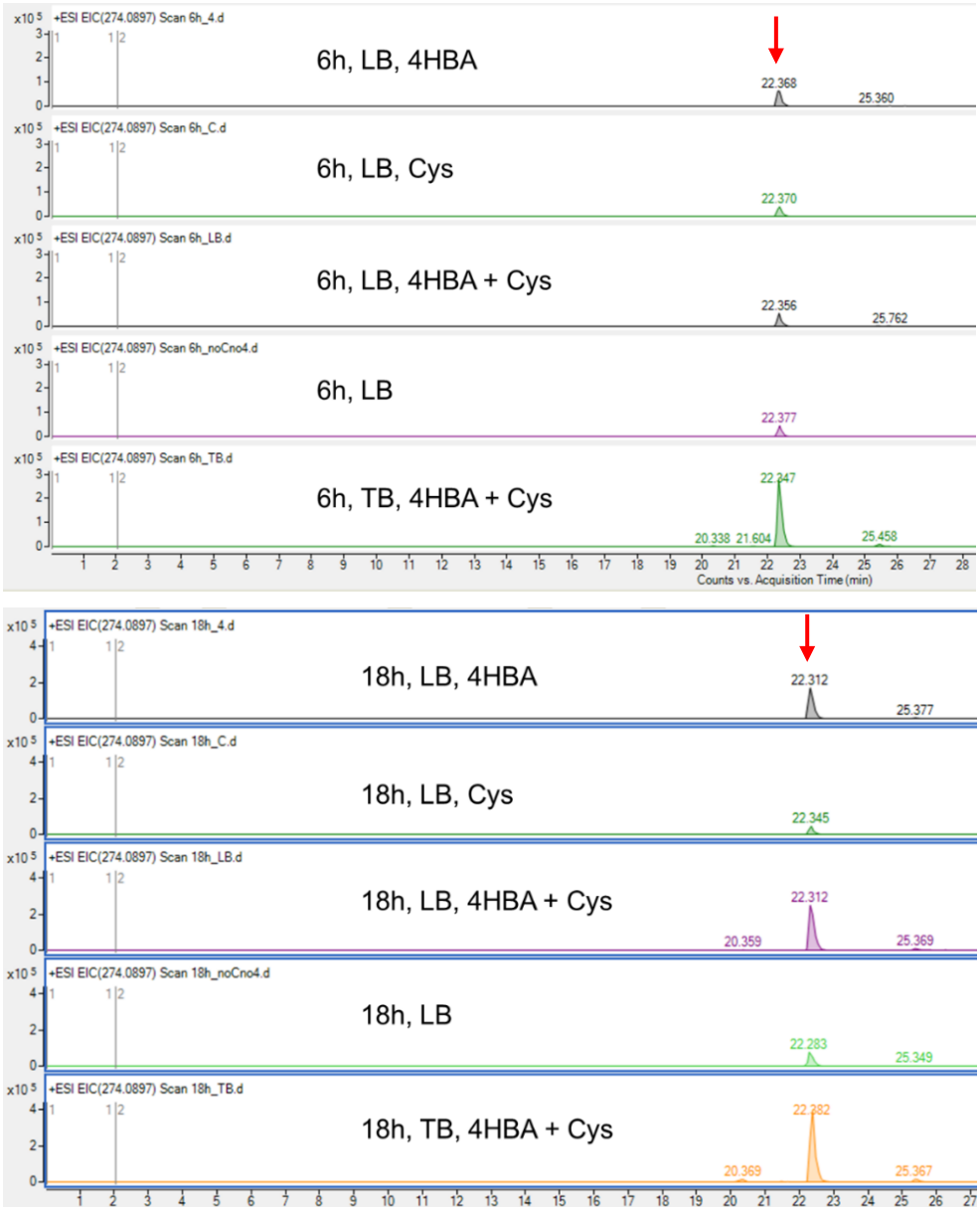
E. Phylogenetic tree of thioesterases, which shows that SarH_TE belongs to a distinct clade with a farther evolutional relationship with PKS/NRPS TEs (TE16-18).

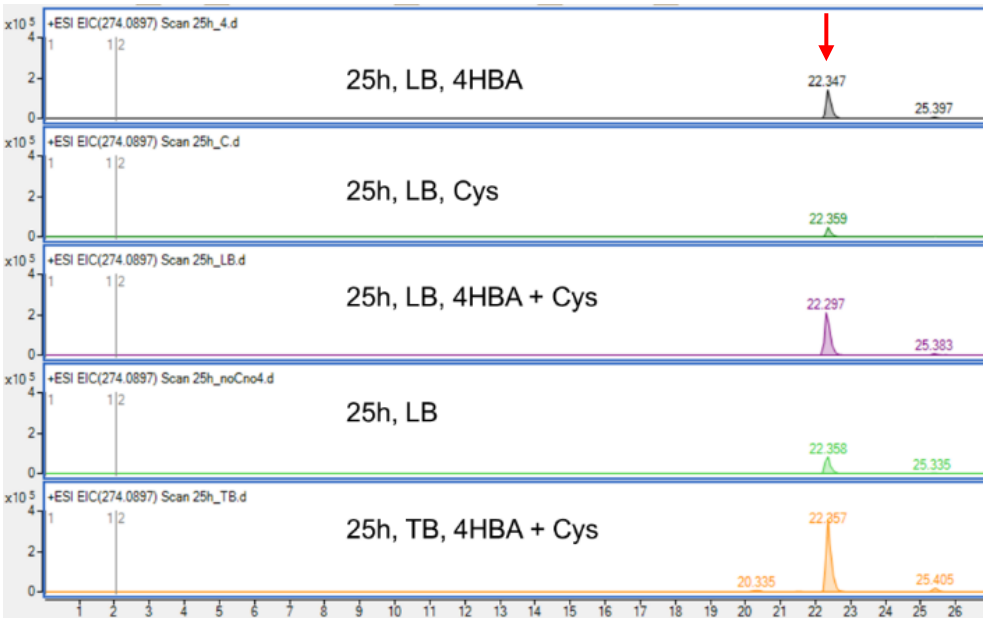


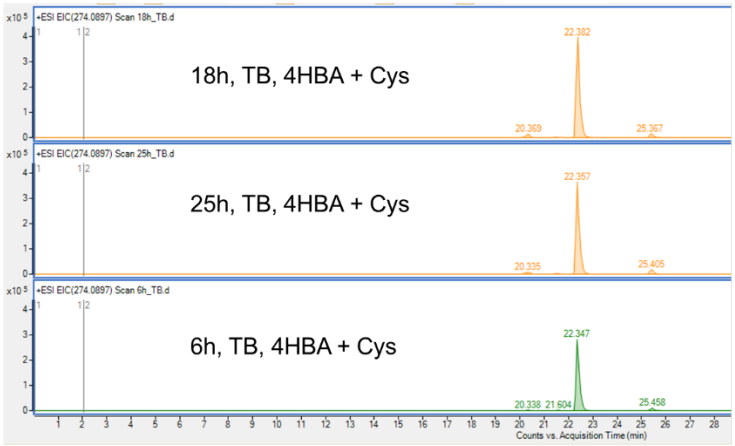
F. Comparison of salivabactin titers between strain **1** in F1 medium and SAL in BHI medium. LC/HRMS measurement with 500 µL culture equivalent injection.



G. Comparison of salivabactin titers from strain **1** under the conditions of different media, growth time and feeding materials. LC/HRMS measurement with 100 µL culture equivalent injection.

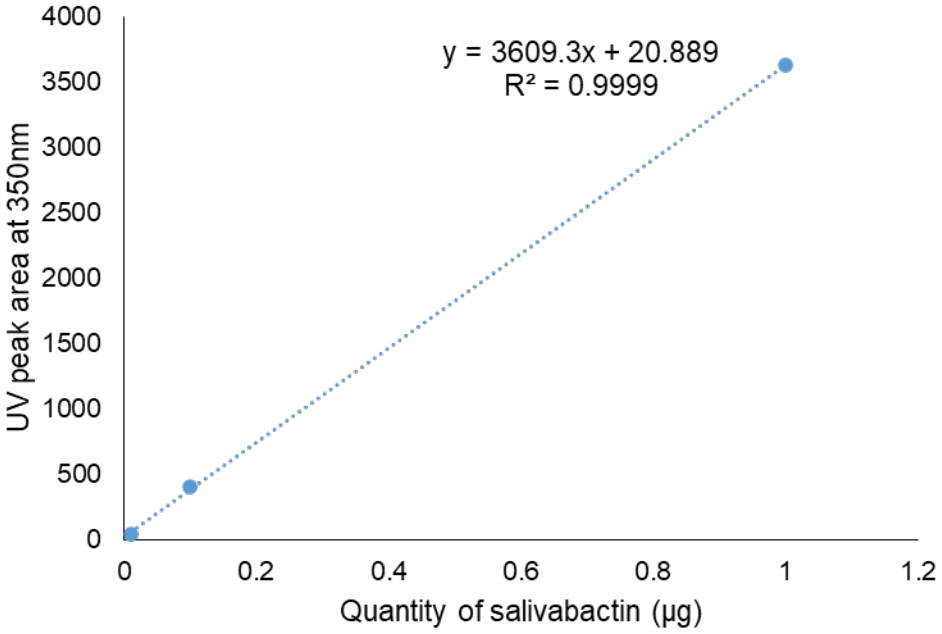




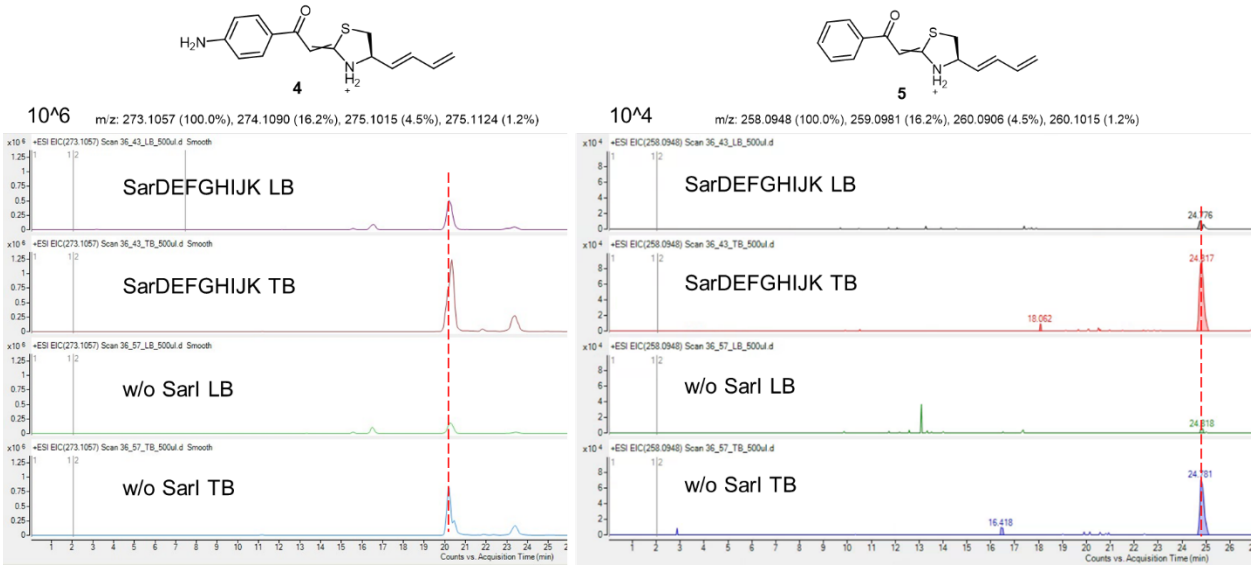


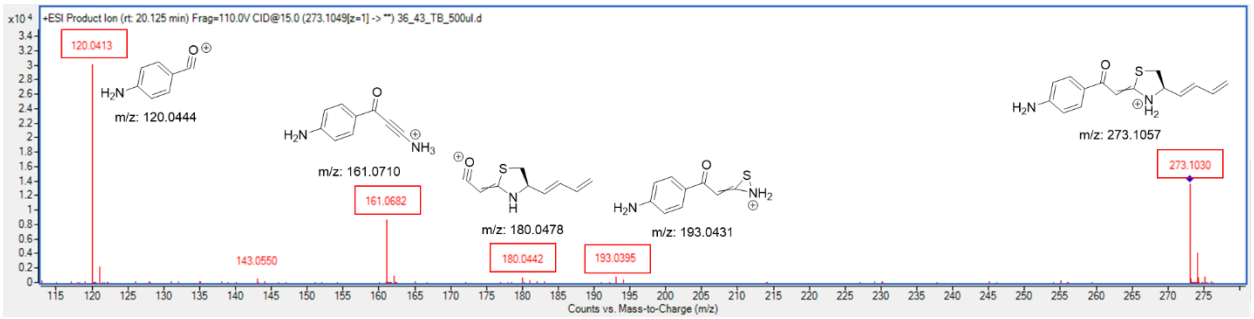
100uL equiv.	MS peak integration area
Strain **1** TB 6h	2,834,004
Strain **1** TB 18h	**4,341,726**
Strain **1** TB 25h	3,698,006
SAL BHI DI	3,320,132
SAL BHI Tap	3,419,753

H. The standard curve correlating the integrated UV area to the quantity of salivabactin, measured by LC-UV/vis-MS.

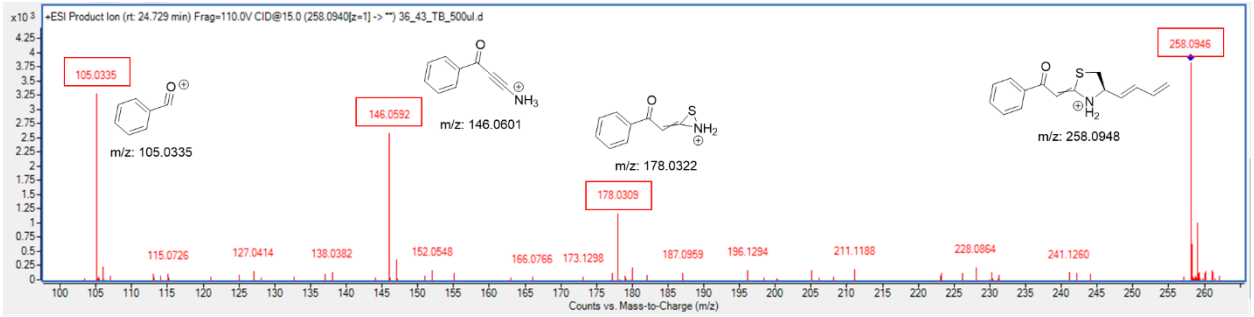


I. Production of compound **4**, **5** and methylated salivabactin in *E. coli*. with the support of MS/MS fragmentation patterns.

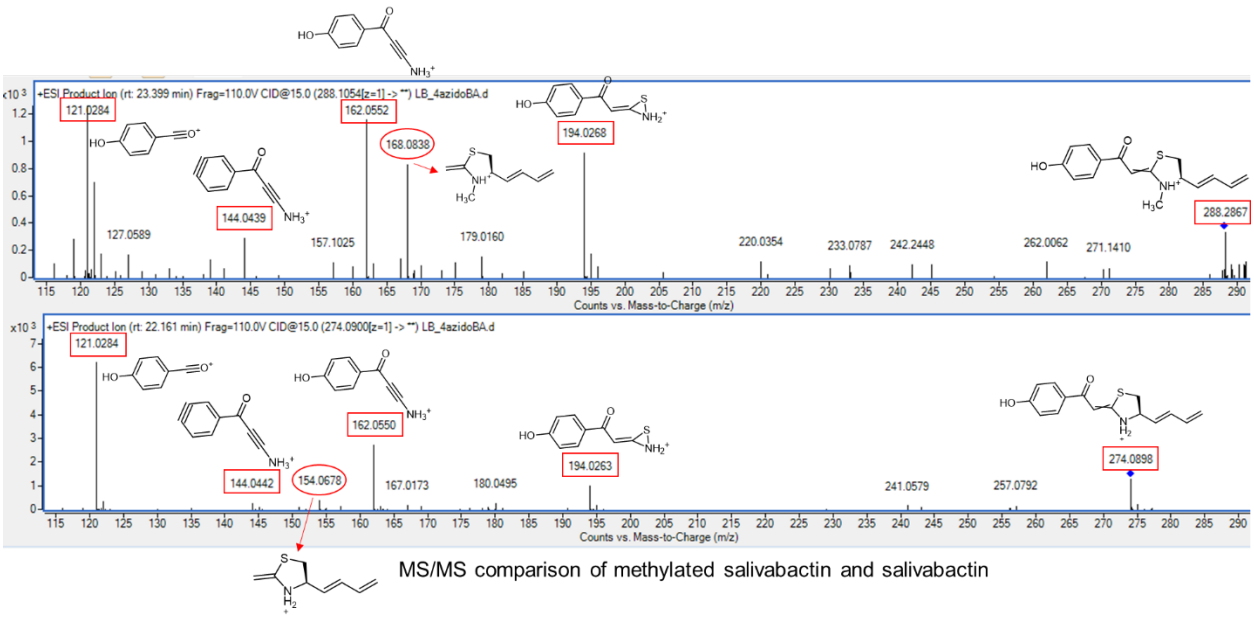




MS/MS fragmentation pattern of compound **4**

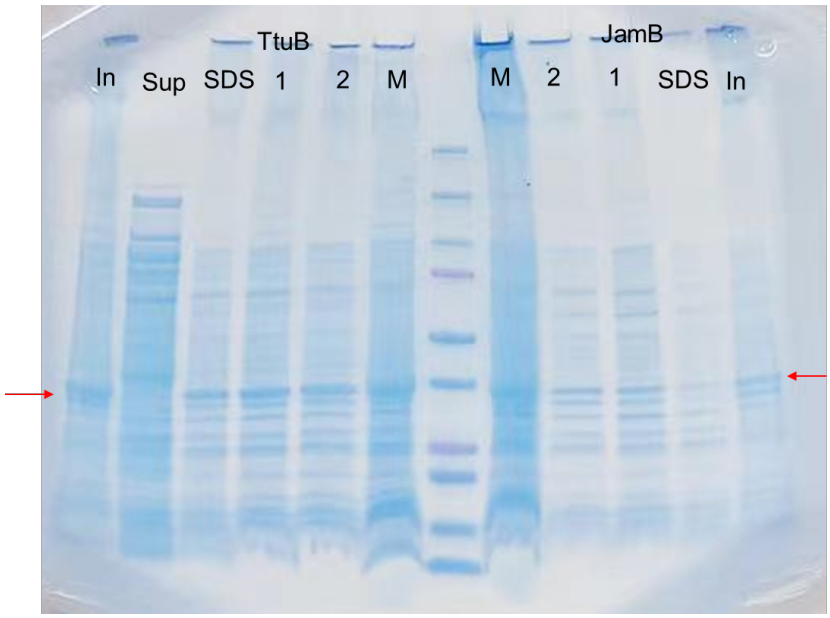


MS/MS fragmentation pattern of compound **5**

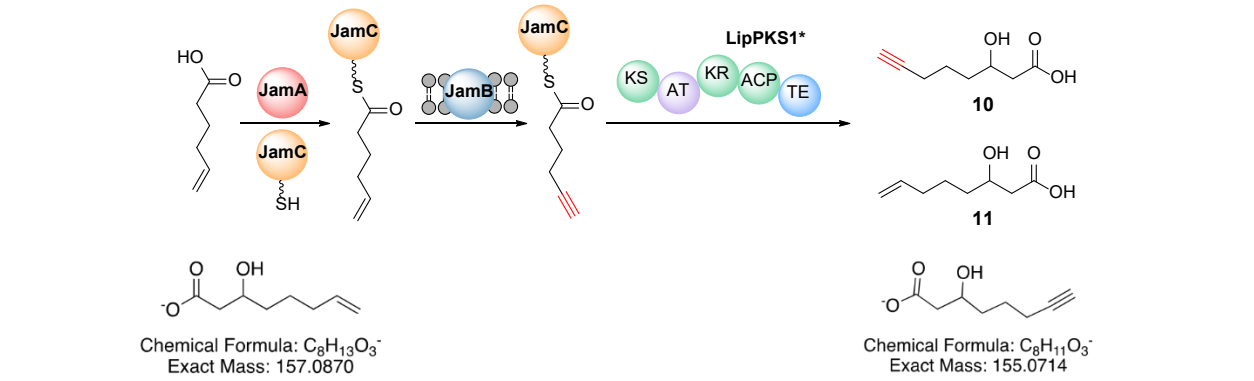


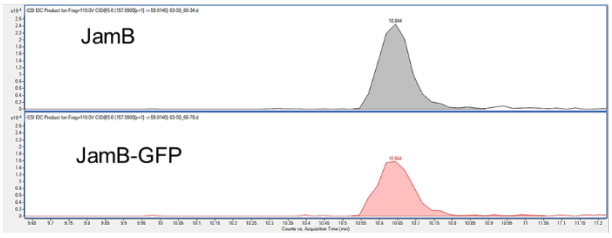
MS/MS comparison of methylated salivabactin and salivabactin

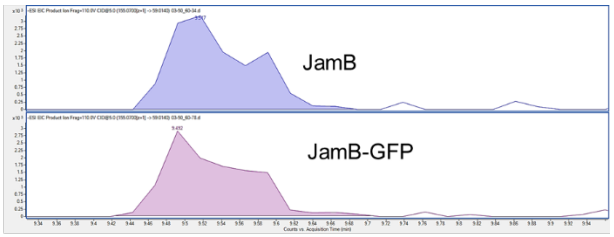
J. Solubilization of JamB and TtuB with detergents. In: insoluble fraction; Sup: supernatant fraction; M: membrane fraction; SDS: sodium dodecyl sulfate; detergent 1: n-Dodecyl-β-D-maltoside; detergent 2: n-Decyl-β-D-maltopyranoside.

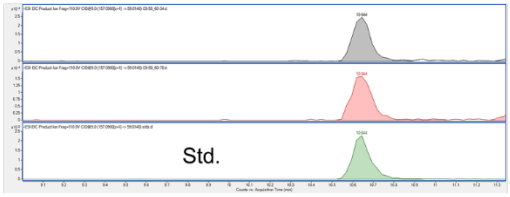


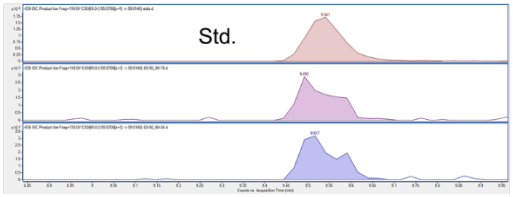
K. Activity tests of JamB-GFP fusion in *E. coli*. Top: biosynthesis and exact masses of compound **10** and **11**. Middle: comparison of activity of JamB and JamB-GFP fusion *in vivo*. The integral areas of MS peaks showed that fusing GFP to JamB does not affect the activity of JamB. Bottom: comparison of the results of JamB, JamB-GFP with the standards of compound **10** and **11**. It showed that JamB and JamB-GFP did produce terminal alkene and terminal alkyne containing natural products.



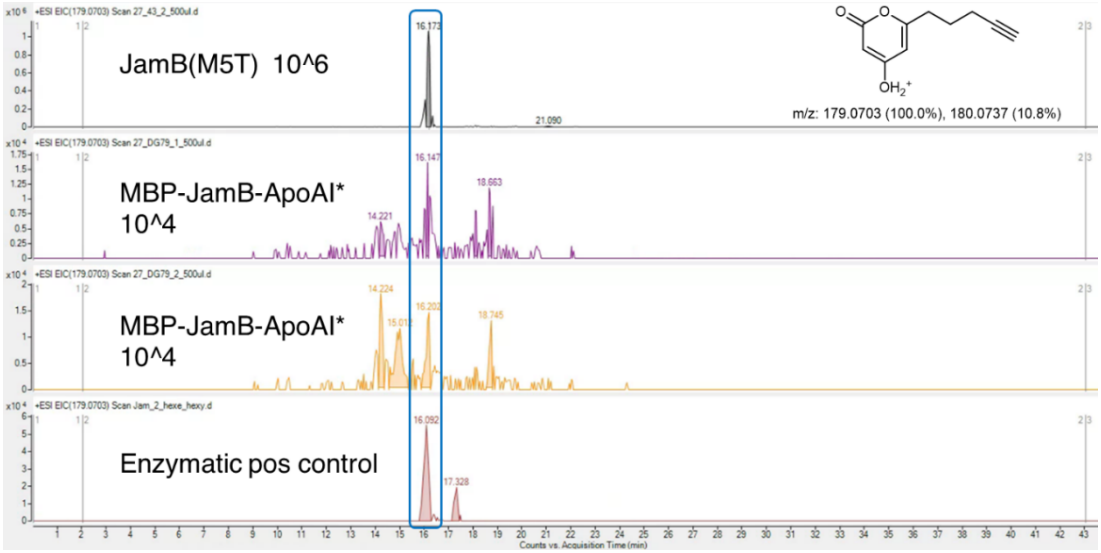


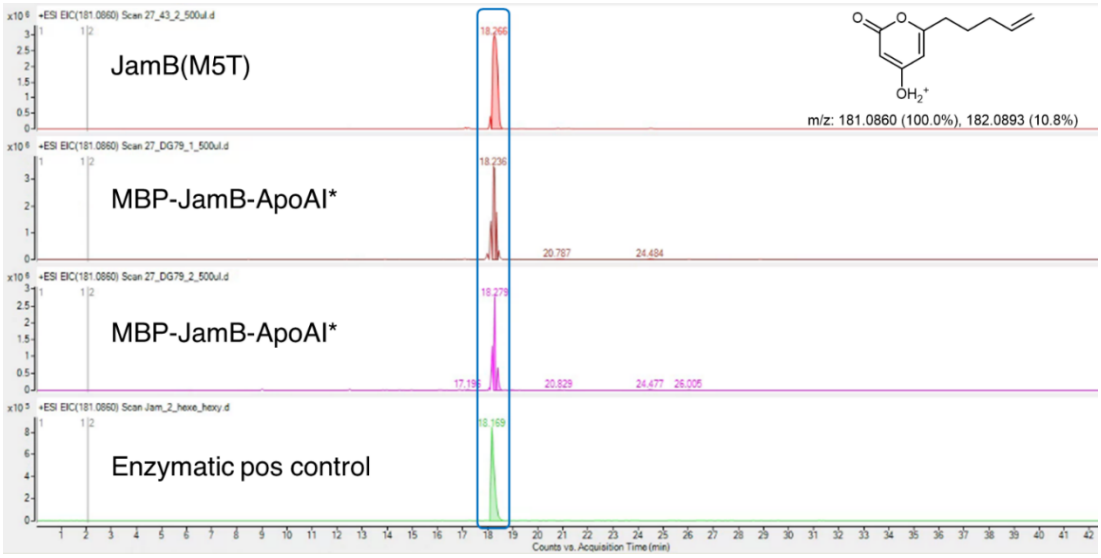




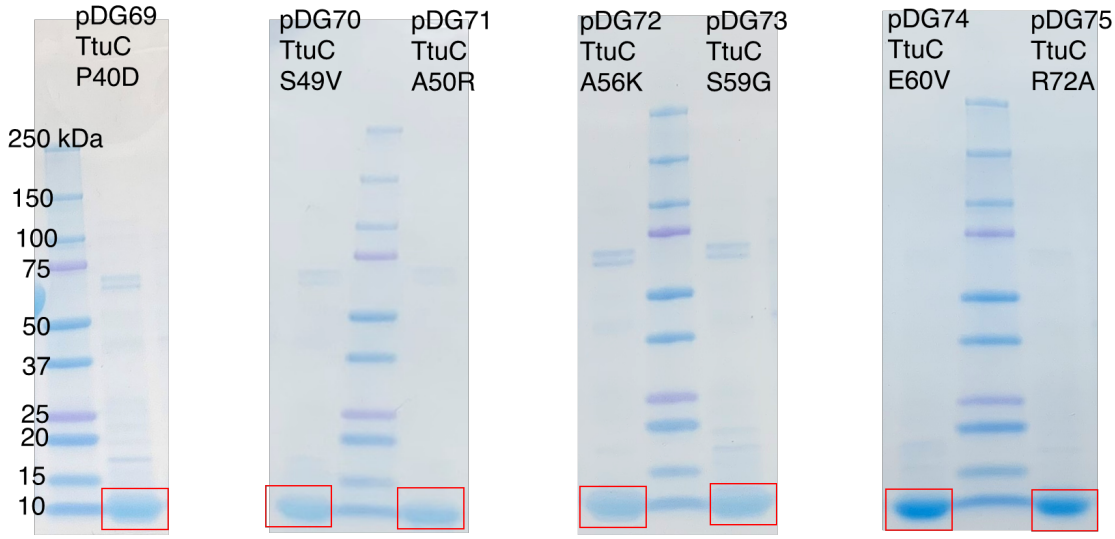


L. Activity tests of JamB engineered with the SIMPLEx method in *E. coli* in F1 Medium, tested in duplicates and measured by LC/HRMS with 500 µL culture equivalent injection.

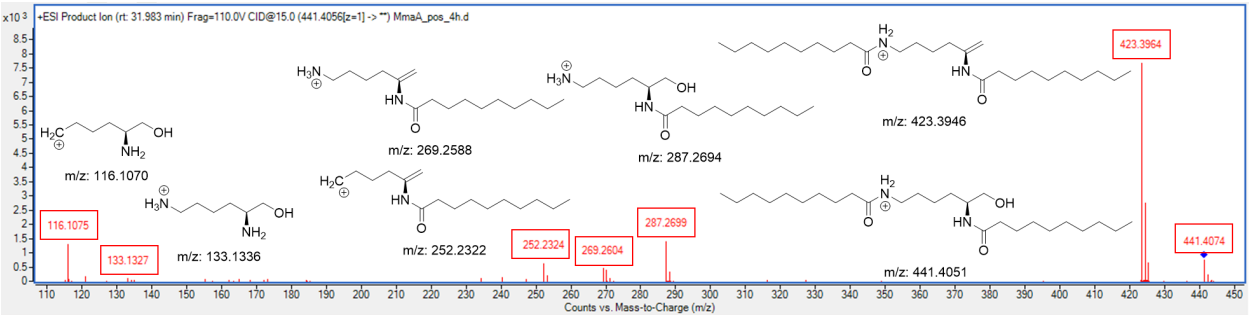




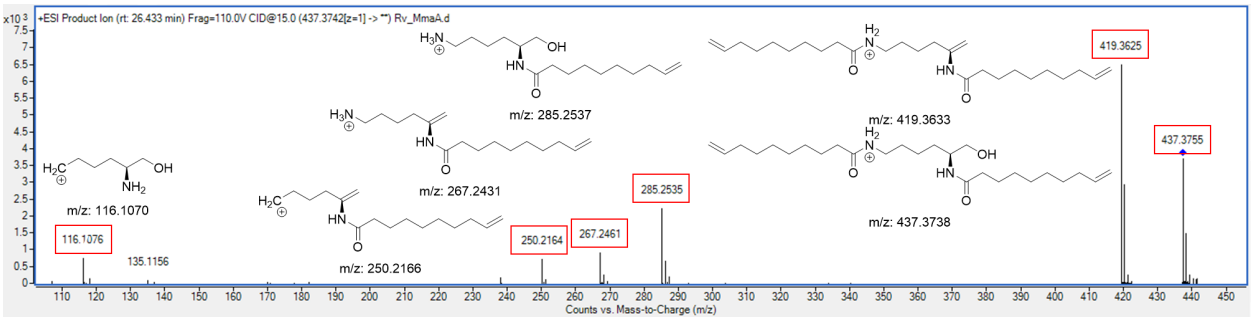
M. SDS-PAGE gels of TtuC mutants.



N. MS/MS fragmentation patterns of compound **15** and **16**.

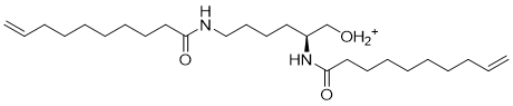


MS/MS fragmentation pattern of compound **15**



MS/MS fragmentation pattern of compound **16**

O. Enzymatic assays to probe the role of the KS^0 domain in the MmaA system. The shift of retention time was due to the maintenance of LC part.



m/z: 437.3738 (100.0%), 438.3772 (28.1%), 439.3805 (3.8%)

