Biosynthesis and heterologous production of vioprolides from myxobacteria

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

Der Universität des Saarlandes

von

Fu Yan

Saarbrücken

2016

Tag des Kolloquiums: 17.03.2017

Dekan: Prof. Dr. rer. nat. Guido Kickelbick

Berichterstatter: Prof. Dr. Rolf Müller

                       Prof. Dr. Jörn E. Walter

Vorsitz: Prof. Dr. Rolf W. Hartmann

Akad. Mitarbeiter: Dr. Maksym Myronovskyi

ACKNOWLEDGEMENT

First of all, I would like to express my sincere gratitude to Prof. Dr. Rolf Müller for offering me the opportunity to work in this high-rank team in Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) and supervising me in the past four years. Thanks to his various kinds of supports in my research, I have not only finished the project but also learned so much during the process. The influence from his enthusiasm and attitude in scientific research will benefit my future career and my life.

I express my special gratitude to Prof. Dr. Youming Zhang from the State Key Laboratory of Microbial Technology at Shandong University in China for his guidance in my research and for his help in my life in these years. Without his support I may not have got the chance to study in Germany.

I especially thank Prof. Liqiu Xia from Hunan Normal University for his support and encouragement in my research and my life in these years. He also supported me in the application of the scholarship.

I owe my special gratitude to Dr. Silke Wenzel for her co-supervision and practical suggestions in projects. I learned a great deal from her scientific attitude and interesting ideas. She also helped in proof-reading the thesis.

Meanwhile, I particularly thank Prof. Dr. Xiaoying Bian from the State Key Laboratory of Microbial Technology at Shandong University in China for his supervision in my research and his help in my life in Germany.

I would like to thank Prof. Dr. Jörn Walter from the Department of Genetics at Saarland University for being my second supervisor and for his suggestions in my research.

I would like to thank Prof. Dr. Uli Kazmaier from the Department of Organic Chemistry at Saarland University for accepting me to be a member of the Graduate school on Natural product research. I benefited a lot from lectures and symposiums.

I extremely thank Dr. Mingli Yan from Hunan University of Science and Technology in China. Ten years ago I joined in his project and learned basic skills in molecular cloning which I am still using today. His guidance inspired my interests in scientific research.

My special thanks to all of my colleagues in Prof. Müller's group for their supports and helps in the last four years. I thank Dr. Liujie Huo for his help in my research and my life in Germany; Dr. David Auerbach for the help in vioprolide project and Christian Burgard in myxochromide project; Ram Awal for the translation of the abstract; Eva Luxenburger, Dr. Thomas Hoffmann and Michael Hoffmann for LC-MS measurement; Dr. Khai Bui, Asfandyar Sikandar, Sebastian Adam and Alexandar von Tesmar for their helps in protein

<div align="right">

Fu Yan

Saarbrücken, 17th August, 2016

</div>

PUBLICATIONS

Yin Jia, Hoffmann Michael, Bian Xiaoying, Tu Qiang, **Yan Fu**, Xia Liqiu, Ding Xuezhi, Stewart A. Francis, Müller Rolf, Fu Jun, Zhang Youming (2015) Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Scientific Reports*, 5, 15081

Bian Xiaoying, Plaza Alberto, **Yan Fu**, Zhang Youming, Müller Rolf (2015) Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide. *Biotechnology and Bioengineering*, 112 (7): 1343-1353

Yan Fu, Auerbach David, Chai Yi, Keller Lena, Tu Qiang, Zhang Youming, Müller Rolf. Biosynthesis and heterologous expression of the vioprolide gene cluster revealing a C domain catalyzed glycerate esterification and a post-assembly maturation. (To be submitted)

CONFERENCE CONTRIBUTIONS

Yan Fu, Chai Yi, Bian Xiaoying, Tu Qiang, Zhang Youming, Müller Rolf (October 2014). Heterologous expression of vioprolides, antifungal and cytotoxic cyclic peptolides from *Cystobacter violaceus*. VAAM Workshop: Biology of Natural Products Producing Microorganisms, Dresden, Germany. (Poster presentation)

Yan Fu, Chai Yi, Bian Xiaoying, Tu Qiang, Zhang Youming, Müller Rolf (July 2014). Biosynthetic mechanism and heterologous expression of vioprolides. The 3rd Summer Symposium of Graduate School of Natural Product Research. (Oral presentation)

Yan Fu, Auerbach David, Bian Xiaoying, Chai Yi, Keller Lena, Tu Qiang, Zhang Youming, Müller Rolf (September 2016). Biosynthesis and heterologous production of vioprolides. VAAM Workshop: Biology of Bacteria Producing Natural Products, Freiburg, Germany. (Oral presentation)

ABBREVIATIONS

A	Adenylation domain
ACP	Acyl carrier protein
AMP	Adenosine monophosphate
Amp	Ampicillin
antiSMASH	antibiotics & Secondary Metabolite Analysis SHell
AT	Acyltransferase domain
ATP	Adenosine triphosphate
bp	Base pair
BLAST	Basic Local Alignment Search Tool
C	Carbon
C domain	Condensation domain
CLF	Chain length factor
Cm	Chloramphenicol
CoA	Coenzyme A
CP	Carrier protein
CV	Column volume
Cy	Cyclization domain
Da	Dalton
DNA	Desoxyribonucleic acid
DTT	1,4-Dithiothreitol
E	Epimerization domain
EBFC	Ester bond forming condensation domain
EDTA	Ethylenediamine tetraacetic acid
EIC	Extracted ion chromatogram
FAAL	Fatty acyl AMP-ligase
FACL	Fatty acyl CoA-ligase
FAS	Fatty acid synthase
Genta	Gentamycin
HPLC	High performance liquid chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRMS	High resolution mass spectrometry
IC_{50}	The half maximal inhibitory concentration
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Km	Kanamycin
kb	Kilo base pairs
KS	Ketosynthase domain
kV	Kilo volt
L	Liter
LB	Luria Bertani
M	Molar
Mg	Magnesium

IV

Abbreviations

mg	Milligram
min	Minutes
ml	Milliliter
mM	Millimolar
MOA	Mode of action
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass-to-charge-ratio
N	Nitrogen
nM	Nanomolar
NMT	N-methyl transferase
NRPS	Nonribosomal peptide synthetase
O	Oxygen
OD	Optical density
Orf	Open reading frame
oxytet	Oxytetracycline
oriT	Origin of transfer
SDS-PAGE	Sodium dodecyl sulfate-Polyacrylamid gelelectrophoresis
TE	Thioesterase domain
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
Pip	Pipecolic acid
PKS	Polyketide synthase
Ppant	Phospopantetheine
RNA	Ribonucleic acid
U	Unit
UPLC	Ultra-high performance liquid chromatography
µ	micro
v/v	Volume to volume
w/v	Weight to volume
Zeo	Zeocin
^{o}C	degree Celcius

VI

ABSTRACT

Natural products from bacteria play an important role in disease control. In this thesis, we investigated the biosynthetic mechanism of vioprolides. Vioprolides are peptolides produced by the myxobacterium *Cystobacter violaceus* Cb vi35 with prominent cytotoxicity and anti-fungal activity. Here we elucidated the biosynthetic pathway of vioprolides and expressed the *vio* gene cluster successfully in heterologous hosts *Myxococcous xanthus* DK1622 (*ΔmchA*), *Burkholderia sp.* DSM7029 and *Pseudomonas putida* KT2440. Vioprolides are initially synthesized as previoprolides which undergo post-assembly deacylation and extracellular secretion. We identified an unusual C domain in the NRPS machinery that catalyzes glycerate esterification in the initiation of the assembly line, and validated this process by *in vitro* biochemical characterization. Furthermore, site-directed mutagenesis was carried out to generate various novel vioprolide derivatives.

ZUSAMMENFASSUNG

Naturprodukte aus Bakterien spielen eine wichtige Rolle bei der Krankheitskontrolle. In dieser Arbeit untersuchten wir die Biosynthese von myxobakteriellen Wirkstoffen, den Vioproliden. Vioprolide sind Peptolide und werden vom Myxobakterium *Cystobacter violaceus* Cb vi35 mit prominenten Zytotoxizität und antifungaler Aktivität produziert. Hier erläutern wir den Biosyntheseweg der Vioprolide. Wir exprimierten das „vio" Gencluster erfolgreich in folgenden heterologen Wirten: *Myxococcous xanthus* DK1622 (*ΔmchA*), *Burkholderia* DSM 7029 und *Pseudomonas putida* KT2440. Vioprolide werden zunächst als Pre-vioprolide synthetisiert, die „Post-Assembly" einer Deacylierung und extrazellulärer Sekretion unterliegen. Wir identifizierten eine ungewöhnliche C-Domäne, die eine Glyzerinveresterung zu Beginn der Biosynthesekette katalysiert und validierten diesen Prozess durch *in vitro* Studien für die biochemische Charakterisierung. Weiterhin wurde „Site-directed" Mutagenese durchgeführt, um verschiedene Derivative der Vioprolide zu erzeugen.

TABLE OF CONTENTS

Table of Contents

1. INTRODUCTION

1.1 Natural products in drug discovery

The use of natural products to treat human diseases dates back to the early recorded history [1]. They played and continue to play notable roles clinically. Natural products from bacteria, fungi cyanobacteria and plants exhibit diverse biological activities such as anti-infectives, cytotoxins, immunosuppressors and antidiabetics etc. [2–4] (Fig. 1.1). Among all of the drugs applied in disease treatment in the last several decades, more than 48% derive from natural products or derivatives thereof. Furthermore, half of the chemically synthesized drugs are based on natural product cores or their mimics [3]. From the 326 anti-infective drugs approved from 1981 to 2014, around 50% of them are natural products-related (Fig. 1.2) [3]. More specially, around 60% of antibacterial agents derive from natural products, and most of them originate from microbes. Comparably, in 174 anticancer drugs in this 34 years period, only 13% are originally synthetic.

Epothilone B
anticancer
Sorangium cellulosum

Rapamycin
Immunosuppressant
Streptomyces hygroscopius

Penicillin G
antibacterial
Penicillium notatum

Artemisinin
antimalarial
Artemisia annua L.

Lovastatin
chlesterol lowing
Aspergillus terrus

Daptomycin
antibacterial
Streptomyces roseosporus
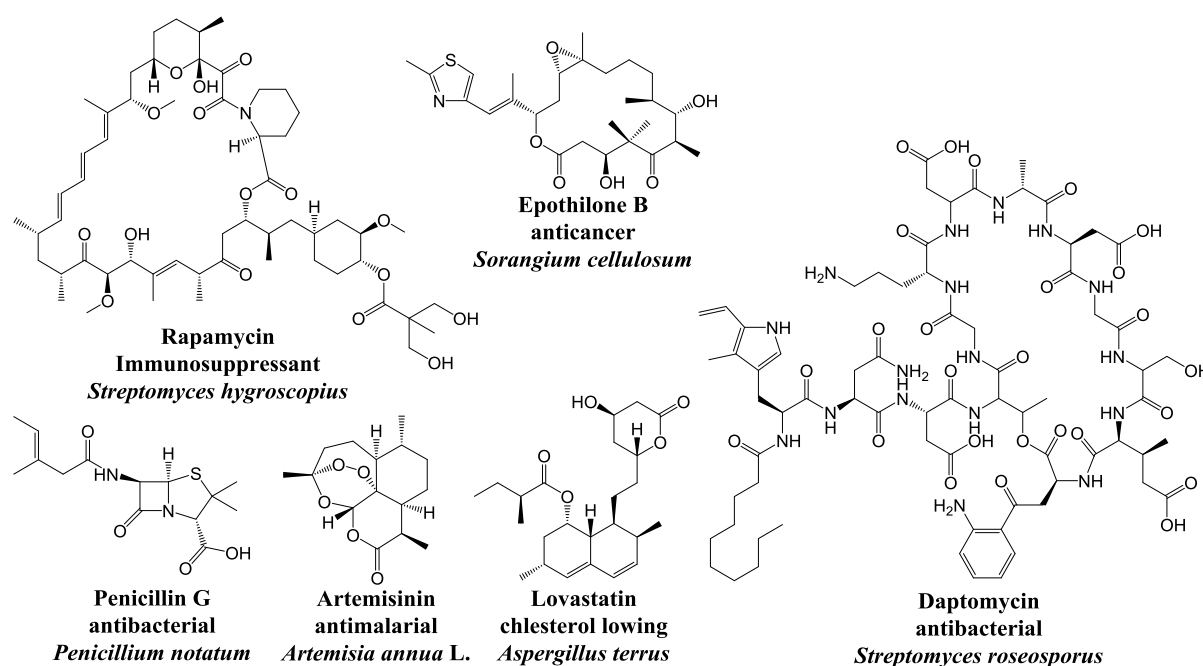
Fig. 1.1 Selected clinically introduced natural products. The name, bioactivity and original producer is shown below the chemical structures.
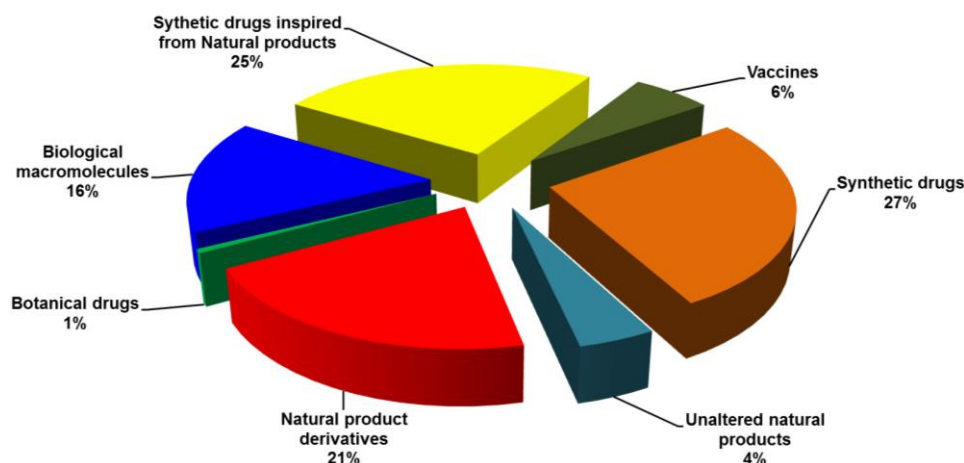
Fig. 1.2 Approved drugs from 1981-2014 (modified from [3]).

Antibiotics have played an important role to cure infectious diseases for more than 70 years. Alexander Fleming's discovery of penicillin in 1928 opened the antibiotic age [5]. Afterwards the development of antibiotics ushered in its golden age. Many important antibiotics which we still use today, e.g. cephalosporins, chloramphenicol, erythromycin, kanamycin and oxytetracycline, were discovered at that time. However, since 1970s the number of newly discovered antibiotics began to decline. Meanwhile, more and more serious antibiotics resistance emerged [6] (Fig. 1.3). Today there are still millions of deaths in the world wide each year caused by the infection of bacterial, fungal, viral or parasitic pathogens. As the widespread misuse of antibiotics increased, accelerating resistance emerged and is now found wide spread in microorganisms [7]. The clinically most relevant ESKAPE pathogens (***E****nterococcus faecium, **S**taphylococcus aureus, **K**lebsiella pneumoniae, **A**cinetobacter baumannii, **P**seudomonas aeruginosa and **E**nterobacter spp*.) are of increasing importance for antimicrobial therapy in the foreseeable future [8–11]. Besides of bacteria infection, infection of fungi *Candidiasis* is common in the world wide. Especially invasive *Candida* usually causes high morbidity and mortality rates [12, 13]. However, only three classes of antifungal agents, the azoles, the echinocandins and the polyenes, are used clinically. Resistance to the

1. Introduction

azoles is becoming common and resistance to the newest antifungal drug echinocandins is also emerging in *Candidiasis* [8]. In contrast to the significant increase of antimicrobial resistance (AMR) in the post-antibiotic-era, nonetheless, pharmaceutical investment in antibiotic development diminished for various reasons and we are facing a severe situation of no usable antibiotics at hand regarding AMR [14]. In the last 40 years only two novel antibiotics, linezolid and lipopeptide daptomycin, were clinically approved [15]. Notably, just several years after introducing linezolid and daptomycin into clinics, resistance in *Staphylococcus* and *Enterococcus* emerged [16–18]. New antibiotics are thus urgently needed, and a new golden age in antibiotic research is probably to solve the current problem.
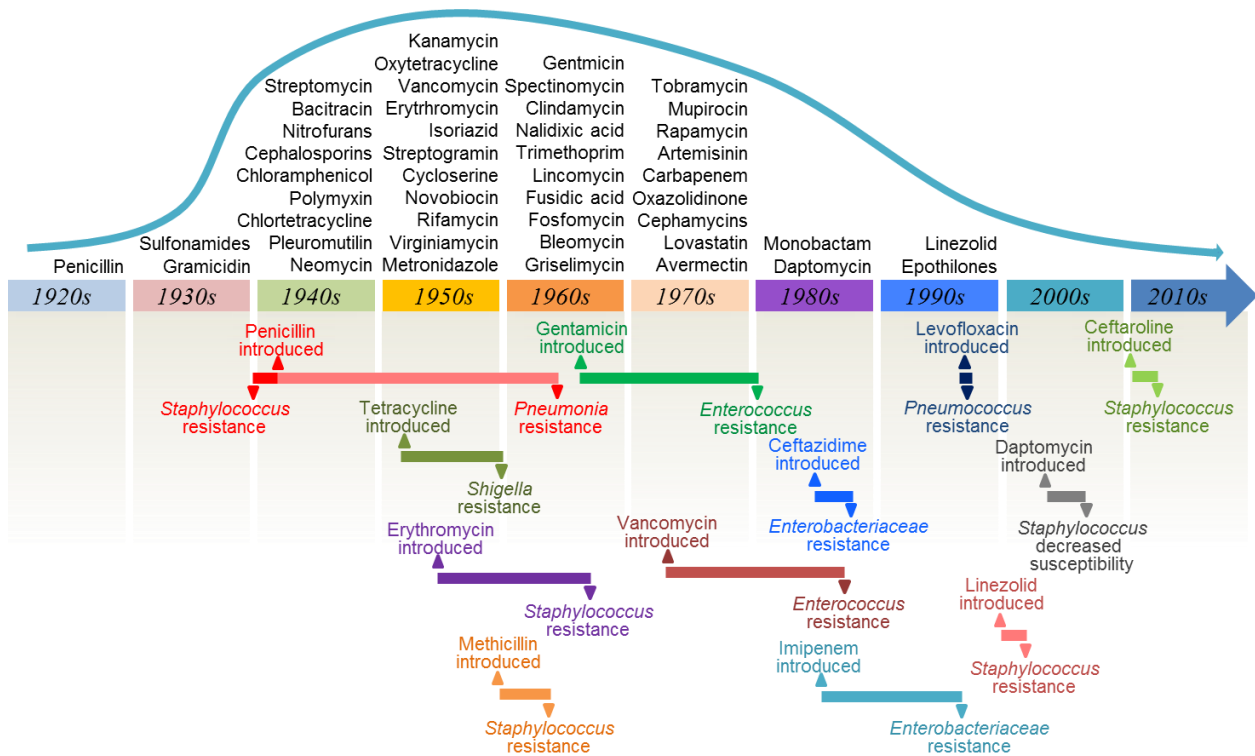
Fig. 1.3 The race between antibiotic development and bacteria resistance. The upper side of the figure shows the antibiotics discovered in each decade; the bars and the arrow in the middle indicate the timeline (decades); the bars below the decades show the time interval from the introduction of antibiotics on the market (upward arrows) to the appearance of resistance (downward arrows). The figure is modified and updated from the reference [11].

1. Introduction

Similar to the severe situation of bacterial infection, cancer is another leading cause of mortality. According to the World cancer report 2014 from the World Health Organization (WHO) [19], there were 14 million new cases with 8.2 million cancer-related deaths in 2012. Furthermore, the annual cases are expected to increase to 25 million in two decades. For a long time, cancer treatment relies on surgery coupled with chemotherapy or radiation therapy. Although these traditional treatments increased survival rates, they could cause side effects such as nausea, diarrhea or even more serious symptoms [20]. Since the first tumor targeting drug rituximab has been approved by the U. S. Food and Drug Administration (FDA) [21], tumor targeting drugs especially small-molecule drugs significantly changed cancer treatment [22–26]. For example, imatinib (also known as Gleevec, Glivec or STI571) is a lankmark of tumor targeting drug, which is a tyrosine-kinase inhibitor approved by FDA in 2001 to treat multiple cancers [27, 28]. An ideal anti-cancer agent should possess high specificity on certain type of cancer cells with low or physiologically acceptable adverse effects. Small molecules from bacteria have great potential to be developed as anticancer leading drugs for their low molecular weight and high cytotoxicity. For instance, the small molecule epithilone analog ixapebilone was approved by FDA to treat metastatic breast cancer, and several of its derivatives are currently in clinical trials [29]. In the condition that a small molecule is highly toxic to both cancer cells and healthy cells, linking the cell-killing small molecule with selective monoclonal antibody for tumor-associated antigen could be a strategy to specifically target the tumor cells [30]. As the antibody-drug conjugates (ADCs) bind to specific tumor antigen and deliver the conjugated active drug inside tumor cells, the anti-tumor effect could be maximized and the damage to healthy tissues could be minimized [31]. Recently, two ADCs, Brentuximab vedotin and Trastuzumab emtansine, have been marketed and several others are in late-phase clinical trial [32, 33].

4

1.2 Biosynthesis of microbial secondary metabolites

Microbial metabolites are traditionally divided into primary and secondary metabolites. Primary metabolites such as citric acid, amino acids and fatty acids are involved in growth, development and reproduction of organisms. Secondary metabolites on the other hand are believed to have important ecological functions including being competitive weapons, agents of symbiosis and metal transporting agents [34]. Millions of years of evolution in microbes generated a myriad of natural products with a variety of structures and bioactivities. Based on their structures and biosynthetic origin, secondary metabolites could be classified into polyketides (PKs), nonribosomal peptides (NRPs), ribosomally synthesized and post-translationally modified peptides (RiPPs), alkaloids, terpenoids, glycosides and others [35]. Up to now, the majority of the characterized microbial secondary metabolites are PKs, NRPs or hybrids thereof which are synthesized by giant enzymes termed polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs) or hybrid PKS/NRPS systems, respectively [36]. This includes clinically used pharmaceuticals, for example, erythromycin A (antibacterial), rapamycin (immunosuppressive), avermectin (anti-parasitic), lovastatin (cholesterol-lowering) and doxorubicin (anticancer) produced by PKSs, as well as cyclosporine (immunosuppressive), tyrocidine (antibacterial) and daptomycin (antibacterial) generated by NRPSs. These giant enzymes recruit monomers such as simple carboxylic acids (PKSs) and amino acids (NRPSs) from primary metabolism or supplied by pathway-specific enzymes to assemble structurally diverse natural products. Although PKS and NRPS use different types of building blocks, they share a similar biosynthetic logic [37, 38].

1.2.1 Polyketide biosynthesis

The earliest genetic study on bacterial antibiotics biosynthesis started around 1970s when Hopwood was investigating genetics in *Streptomyces coelicolor* and discovered antibiotic biosynthetic and resistance genes on plasmid [39], the antibiotic was confirmed to be methylenomycin A shortly afterwards [40]. Thereafter the biosynthetic study on actinorhodin and tetracenomycin C revealed that their PKSs have similarity with fatty acid synthases (FASs) [41–45]. Inspired from this, Leadlay and Katz performed the landmark investigations on the erythromycin polyketide synthase and dramatically accelerated the study on biosynthetic pathways of natural products [46, 47].

A minimal PKS module consists of an acyltransferase (AT) domain (~50 kDa), a ketosynthase (KS) domain (~45 kDa) and an acyl carrier protein (ACP, 8~10 kDa) [36]. Prior to playing a role, *apo*-ACP domains have to be primed by phosphopantetheinyltransferases (PPTases) [37]. PPTase transfers a thiol-terminated pantetheinyl arm from Coenzyme A (CoA) to the hydroxyl group of the active site serine in *apo*-ACP to generate *holo*-ACP (Fig. 1.4). AT domains specifically transfer C2, C3 or C4 acyl groups from malonyl-, methylmalonyl- or ethylmalonyl-CoA to *holo*-ACP to form acyl-*S*-pantetheinyl-ACP. KS domains catalyze the chain extension by decarboxylative Claisen condensations between monomers and the growing polyketide chain. In addition to the three domains essential to chain elongation (AT, ACP, KS), ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains may also be part of PKSs. DH, KS and ER domains function sequentially. DH domain reduces β-keto group to β-hydroxyl, KR domain dehydrates β-hydroxyl group to α, β-olefin, and ER domain reduces α, β-enyl group to β-methylene [36]. Because of the existence or absence of KR, DH or ER domain, the β-keto group of the extending chain could be processed as β-ketone, β-hydroxyl, α, β-olefin or β-methylene (Fig. 1.5). Finally, when an approximate chain length is achieved, the

acyl chain is released by a thioesterase (TE) domain via hydrolysis or macrocyclization [48]. The released products may undergo further modifications such as glycosylation, hydroxylation or methylation by dedicated enzymes to obtain bioactivities [49]. According to the domain organization and incorporated substrates, PKSs could be characterized into three subgroups: type I, type II and type III PKS. Domains in type I PKSs are connected *in cis*, while in type II PKSs they are discrete and function *in trans*. Instead of using malonyl-*S*-pantetheinyl-ACP as substrates by type I and type II PKSs, type III PKSs directly use manolyl-CoA as substrates [50].

apo-CP   Coenzyme A   *holo*-CP
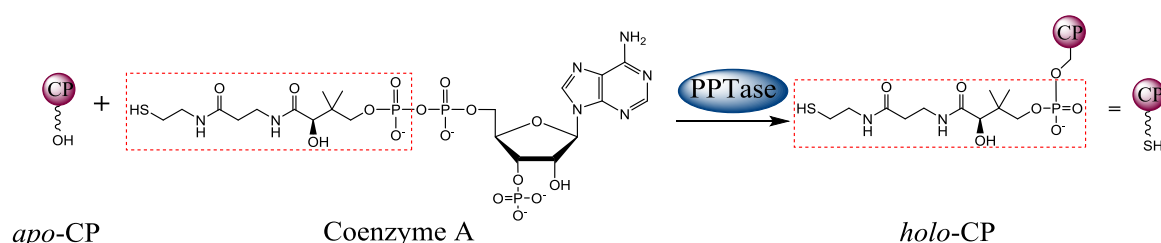
Fig. 1.4 Priming of carrier proteins (CP) by phosphopantetheinyltransferases (PPTase). Phosphopantetheine arm is transferred from coenzyme A to the conserved serine of carrier protein (ACP or PCP domain).
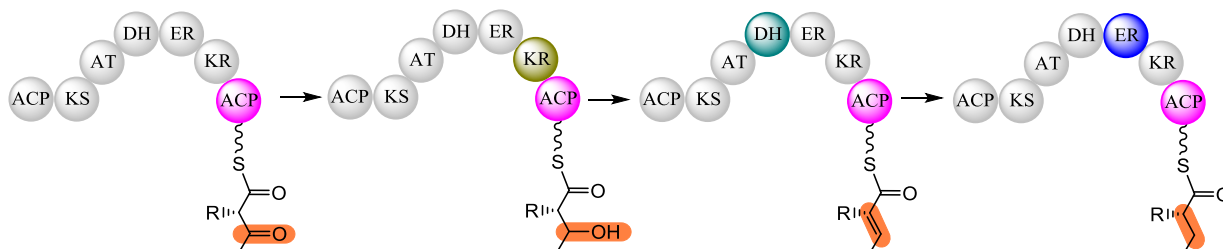
Fig. 1.5 β-keto processing in Type I PKs and FAs. Active domains in each step are shown in colors. The modified β-keto is highlighted.

1.2.2 Nonribosomal peptide biosynthesis

The earliest biochemical investigation on gramicidin S biosynthetic enzymes by Lipmann revealed a biosynthetic route which was different from ribosomal-linked protein synthesis, and amino acids were supposed to be activated as

aminoacyl-AMP and transferred to carrier protein [51]. The following research unveiled that peptidyl transfer between intermediates was bound to thioester linkages of two gramicidin S synthetases [52]. Shortly afterwards peptides were discovered attached to pantetheine by a thioester bond in gramicidin S and tyrocidine biosynthesis [53, 54]. Almost at the same time the biosynthetic genes of the siderophore enterobactin (also known as enterochelin) were identified [55–57]. After that the biosynthetic genes of tyrocidine, gramicidin S and surfactin were also identified by Marahiel [58–62]. The biochemical investigations on enterobactin biosynthesis by Walsh characterized PPTase and demonstrated the formation of adenylated substrates, and the biosynthetic model of enterobactin was finally uncovered [63–73].

The biosynthetic logic from NRPSs is in principle similar to modular PKSs, although they apply a different biochemistry for recruitment and condensation of substrates. NRPSs use amino acids rather than acyl-CoA as building blocks. The core domains of NRPSs are adenylation (A) domain (~50 kDa), peptidyl carrier protein (PCP, ~8-10 kDa) and condensation (C) domain (~50 kDa) [36, 74]. Similar to PKSs, *apo*-PCPs are modified to *holo*-PCP by PPTase before starting the assembly (Fig. 1.4). Generally, the A domain selects an amino acid and activates it as aminoacyl-adenylate at the expense of ATP and Mg^{2+}. The aminoacyl-AMP is then attacked by thiol moiety of the *holo*-PCP to form aminoacyl-PCP. The C domain plays the critical role in the elongation of the peptide chain. It catalyzes the amide bond (C-N) formation between upstream peptidyl-PCP (electrophilic) and downstream aminoacyl-PCP (nucleophilic) (Fig. 1.6). When the peptide chain reaches up to the last PCP, it is transferred to the TE domain and catalytically released by hydrolysis or intramolecular cyclization [35, 48, 75]. Exceptionally, some optional domains such as epimerization (E) domain, cyclization (Cy) domain, oxidase (Ox) domain, dual condensation/epimerization (C/E) domain, methyltransferase (MT) domain and

aminotransferase (AMT) domain etc. could also exist in NRPSs [36]. The E domain epimerizes L-amino acids to D-amino acids, its downstream C domain is designated $^{D}C_L$ domain for the function to condense D-peptidyl donor with L-aminoacyl acceptor [76]. In some NRPSs, E domain and C domain are replaced by an individual bifunctional C/E domain which is equivalent to E-C didomain [77]. Similar to C/E domains, Cy domains represents also bifunctional domains. Cy domains catalyze amide bond formation between Cys, Ser or Thr and an upstream aminoacyl donor. After the condensation event, a heterocycle is formed via nucleophilic attack of the Cys, Ser or Thr side chain (thiol or hydroxyl group) on the upcoming carbonyl group, followed by elimination of water to form thiazoline or (methyl)oxazoline ring structures [36]. These can be further converted to thiazole or (methyl)oxazole moieties by Ox domain, respectively. Thiazoline could also be reduced to thiazolidine by reductase, as exemplified in pyochelin biosynthesis [78]. NRPSs could also integrate MT domains to transfer methyl group from *S*-adenosylmethionine (SAM) to α-amino group, or AMT domains to transfer α-amino group from glutamine to the β-carbon of acyl chain [35, 79].
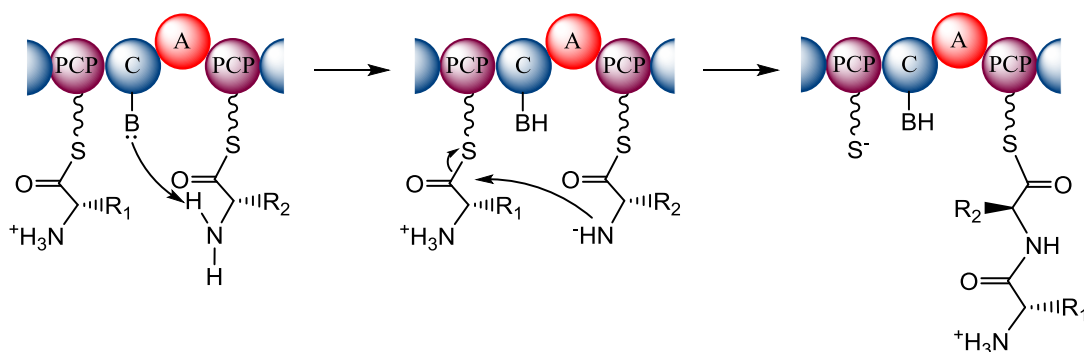
Fig. 1.6 C-N bond formation in NRPs biosynthesis.

1.2.3 Biosynthesis of polyketide-peptide hybrid compounds

Besides of pure PK or NRP structures, a large number of natural products such as epothilone, rapamycin and bleomycin etc. represent polyketide-peptide

hybrids [80]. They are manufactured by assembly lines harboring PKS and NRPS modules, socalled hybrid PKS/NRPS megasynthetases. However, PKS/NRPS systems have a few specific characteristics. Firstly, PPTase should be able to trim both ACPs and PCPs in hybrid systems [81]. Secondly, docking regions of NRPS domain and PKS domain should crosstalk well [82]. Thirdly and most importantly, C domain in PKS/NRPS system should recognize polyketide donor, while KS domain in NRPS/PKS machinery should accept peptide chain [35]. As hybrid machineries harbor the features of both NRPS and PKS, structural diversity can be further increased by combining PKS/NRPS biochemistry.

1.2.4 Lipopeptide biosynthesis

Lipopeptides represent a prominent class of NRPSs and PKS/NRPS hybrid products exhibiting different biological activities such as antimicrobial, antimalarial and anticancer activities [15, 83–85]. They are produced by many bacteria including bacillus, pseudomonads, actinomycetes, cyanobacteria and myxobacteria. Most of the lipopeptides composed of a cyclic core peptide acylated at the N-terminus with a lipid chain, while some others are formed by a linear peptide and 1~3 lipid chains [15, 86, 87] (Table 1.1). Majority of the lipopeptides like surfactin [88–92], daptomycin [93–96] and telomycin [97] etc. are synthesized by NRPSs. While some others eg. cystomanamides [98] and myxochromides [99] are synthesized by PKS/NRPS hybrids. In daptomycin and telomycin biosynthesis, fatty acid was activated by an independent fatty acyl AMP-ligase (FAAL) and loaded to a free-standing ACP. The N-terminal C domain then condenses fatty acid with the first amino acid before assembly of peptide core. The biosynthesis of surfactin and calcium-dependent antibiotic (CDA) is initiated in a similar way, but instead of being carried by an ACP, fatty acid is carried by CoA. In some cases, e.g. as described for xenocoumacin [100], colibactin [101], zwittermicin [102], didemnin [103] and telomycin [97], lipid

chains are introduced only temporarily and are hydrolyzed at a later stage / the end of the biosynthesis.

Table 1.1 The growing class of lipopeptides*

Lipopeptides	Origin	Bioactivity
A21978C	*Streptomyces roseosporus*	Antibacterial
A54145	*Streptomyces fradiae*	Antibacterial
Amphomycin	*Actinoplanes friuliensis*	Antibacterial
Amphisins	*Pseudomonas sp.* DSS73	Antifungal
Arthrofactin	*Pseudomonas sp.* MIS38	Antifungal
Bacillomycins	*Bacillus subtilis*	Antifungal
Carmabin A, Dragomabin, and Dragonamide A	*Lyngbya majuscula*	Antimalarial
Cerexins (**Linear lipopeptide**)	*Bacillus cereus* 60–6 and Gp-3	Antibacterial
Calcium-Dependent Antibiotics	*Streptomyces coelicolor*	Antibacterial
Cystargamide	*Kitasatospora cystarginea*	NB
Cystomanamides (**Linear glycolipopeptide**)	*Cystobacter fuscus* MCy9118	NB
Daptomycin	*Streptomyces roseoporous*	Antibacterial
Echinocandins	*Aspergillus nidulans var echinulatus, Coleophoma empetri F-11899, Zalerion arboricola*	Antifungal
Fengycins	*Bacillus subtilis, Bacillus amyloliquefaciens*	Antifungal
Friulimicin	*Actinoplanes friuliensis*	Antibacterial
Fusaricidins	*Paenibacillus spp.*	Antibacterial and antifungal
Gavaserin	*Bacillus polymyxa*	Antibacterial
Iturins	*Bacillus subtilis, Bacillus amyloliquefaciens*	Antibacterial and antifungal
Jolipeptin (**Linear lipopeptide**)	*Bacillus polymyxa*	ND
Kurstakin A	*Bacillus thuringiensis subsp.* kurstaki HD-1	Antifungal
Lichenysins	*Bacillus licheniformis* IM 1307	Antibacterial
Locillomycins	*Bacillus subtilis*	Antibacterial and antiviral
Marihysin A	*Bacillus marinus* B-9987	Antifungal
Massetolide A	*Pseudomonas fluorescens* R1SS101	Antifungal
Mixirins	*Bacillus sp.*	Anticancer
Mojavensin	*Bacillus mojavensis* B0621A	Antifungal and anticancer
Mycosubtilin	*Bacillus subtilis*	Antifungal
Myxochromides	*Myxococcus sp., Stigmatella sp., Hyalangium minutum*	NB
Octapeptins	*Bacillus and*	Antibacterial and antifungal

	Paenibacillus spp.	
Paenibacterins	*Paenibacillus sp.* OSY-SE	Antibacterial
Pelgipeptins	*Paenibacillus elgii* B69	Antibacterial and antifungal
Plipastatins	*Bacillus cereus* BMG302-FF67	anti-inflammatory
Polymyxins	*Paenibacillus*	Antibacterial
Polypeptins	*Bacillus circulans*	Antibacterial
Pumilacidins	*Bacillus pumilus*	Antiviral
Putisolvin	*Pseudomonas putida* PCL1445	Antibacterial
Puwainaphycin	*Cylindrospermum alatosporum*	Cytotoxic
Ramoplanin (**Glycolipopeptide**)	*Actinoplanes sp.*	Antibacterial
Saltavalin (**Linear lipopeptide**)	*Bacillus polymyxa*	ND
Subtulene	*Bacillus subtilis* SSE4	Antibacterial and antifungal
Surfactins	*Bacillus subtilis*	Antimicrobial, antiviral, antimycoplasma, anticancer and anti-inflammatory
Surotomycin	*Streptomyces roseosporus*	Antibacterial
Syringopeptin	*Pseudomonas syringae pv. Syringae* B728a and B301D	Antibacterial, antifungal and phytotoxic
Syringomycin	*Pseudomonas syringae pv. Syringae* B728a and B301D	Antifungal and phytotoxic
TAN-1511 (**Linear lipopeptide**)	*Streptosporangium amethystogenes*	Anti- leucopenia
Telomycin	*Streptomyces canus* C159	Antibacterial
Tolaasin	*Pseudomonas tolaasii, Pseudomonas sp.* NZ17	Antibacterial, antifungal and phytotoxic
Tridecaptins (**Linear lipopeptide**)	*Paenbacillus polymyxa*	Antibacterial
Variochelin	*Variovorax boronicumulans*	Siderophore
Previoprolides **In this study** (**Lipopeptolide**)	*Cystobacter violaceus* Cb vi35	Anticancer and antifungal
Viscosins	*Pseudomonas libanensis, Pseudomonas fluorescens*	Antibacterial and anticancer

*, the list is summarized from reference [15, 83–87, 97–99, 104–108]. ND, no data; NB, no obvious bioactivity at present.

1.3 Genetic engineering and heterologous expression in NP research

1.3.1 Genetic engineering in natural products research

The total synthesis or semi-synthesis of complicated natural products is usually challenging. Bioengineering approaches provide possibilities to develop modified compounds with improved pharmaceutical properties. In the early

stage of natural product research, precursor directed biosynthesis was used to generate natural product derivatives [109, 110]. A problem with this method is that the modified precursors have to compete with native ones, thus the yield of final derivatives is sometimes not prominent. Mutasynthesis could avoid this problem by inactivating native precursor supplying genes, but the major limitation of this method is the inefficient or unavailable incorporation of synthetic precursors due to the limited substrate flexibility of the native enzymes [109, 111]. Engineering precursor pathways endogenously or introducing tailoring enzymes such as halogenases, glycosylases, acyltransferases and sulfotransferases etc. from other gene clusters represent promising opportunities for biosynthetic engineering and addressing challenges associated with precursor directed biosynthesis and mutasynthesis approaches [109].

The linear order and composition of domains in modular PKS and NRPS determine the structure of the final products. To some extent, it is possible to modify the structure specifically and to produce new compound libraries by re-programming PKS or NRPS machineries via genetic engineering [112]. For example, by exchange *dapD* gene in daptomycin biosynthetic gene cluster with *cdaPS3* (from CDA gene cluster) or *lptD* (from A54145 gene cluster), the last moiety kynurenine in daptomycin structure was changed into Val, Ile or Trp [113]. The same compounds were also obtained by exchanging C-A didomain (kynurenine condensation) in the last module of daptomycin biosynthetic machinery with the respective part of CDA or A54145 megasynthetase [114]. Besides, the binding pocket of A domain contains some specificity-conferring codes, the selectivity of A domain could be altered by changing these codes [115–117]. Take gramicidin S NRPS for example, W239S mutation in $GrsA_A$ changed the selectivity from Phe to unnatural aromatic amino acid *O*-propargyl-L-tyrosine without effects on catalytic efficiency [116]. Furthermore, directed

evolution or random mutagenesis of A domains is also an efficient way to generate new derivatives [109, 118].

Similar modification strategies in NRPS engineering could also be utilized on PKSs. Starter units are utilized only once and probably could be tolerated by downstream modules in PKSs, consequently non-natural starter units might be introduced to PKs by exchange loading modules [119]. Module insertion or deletion is also feasible to generate novel polyketide structures. In addition, promiscuity of AT and KS domain could be used to introduce non-natural extender units. As represented in the engineering of 6-Deoxyerythronolide B (DEBS) PKSs, numerous erythromycin analogs were obtained by exchange methylmalonyl-specific AT domain to AT domains with different substrate specificities [120–122] or by site-directed mutagenesis on AT domain [123, 124]. Moreover, structural diversity could be achieved by manipulation on DH, KR and ER domains. For example, nystatin derivatives with modified antifungal and hemolytic activities were generated by site-specific inactivation of the DH, KR and ER domains in its PKS [125].

1.3.2 Heterologous expression

Despite the fact that genetic engineering methods were applied in many native producers to achieve the production of compounds, genetic manipulation of original hosts is in many cases tedious or even infeasible. Especially, many myxobacteria are hard to cultivate, and genetic tools are not well developed in this kind of bacteria. Although increasing precursor supply, multiplication of gene cluster dosage, manipulating regulators and repressing competing pathways etc. could be used to elevate the production yield, the expression level of many biosynthetic gene clusters in native hosts is low [111, 126]. In addition, whole genome sequencing reveals that a large number of gene clusters in native producers are "cryptic" or "silent", and compounds are undetectable or not produced in laboratory conditions [127]. Despite that these cryptic gene clusters

might be activated by 'physiological triggers' and 'metabolic engineering' [127], the process could be labour-intensive and time-consuming. Moreover, if the native producer is slow growing, improvement on it can be more difficult. In these conditions, a genetically amenable heterologous host has an advantage in characterization of biosynthetic gene clusters, expression of biosynthetic pathways, generation of derivatives and increasing production titers [128]. Nevertheless, there are also some challenges in heterologous expression. For instance, the biosynthetic gene clusters are usually very large. Depending on the characteristics of the bacteria, introducing a large DNA fragment into heterologous hosts could be difficult. In some bacteria, e.g. myxobacteria, plasmids usually could not be sustained in the cells, thus it is necessary to integrate the biosynthetic gene cluster into chromosome. Due to the complexicity of biosynthetic machineries, megasynthetases may not be expressed or correctly folded into functional forms in heterologous hosts. Therefore, the compound production level in heterologous hosts could also be low or even zero [129, 130].

In general, heterologous expression involves five steps: cloning of the biosynthetic gene cluster, engineering of expression vector, transformation of heterologous hosts, fermentation and product detection (Fig. 1.7). Several technologies, such as ligation-dependent conventional cloning and Golden Gate assembly [131], Gibson assembly [132], recombination-mediated Red/ET recombination [133–136] and transformation-associated recombination (TAR) [137, 138] could be utilized to clone the target gene cluster. Based on bioinformatic analysis, large gene clusters could also be directly synthesized and assembled from small fragments. During *de novo* DNA synthesis, codon usage could be optimized and domain organization in the assembly lines could be elaborately designed. Generally, the target gene cluster is cloned to an expression vector and a series of genetic manipulations need to be performed on

the gene cluster as well as the vector backbone. Suitable genetic elements such as resistance genes, replicons or elements for chromosomal integration (fragments for homologous recombination, transposon or phage integrase cassettes) may have to be added to maintain the gene cluster in heterologous hosts. Usually the native promoter of the gene cluster has to be replaced by inducible or constitutive promoter to produce target compounds in high titer [129]. Sometimes tailoring enzyme encoding and precursor supplying genes also need to be added to the expression vector. When the object goes to generate derivatives or gene knock-out, mutagenesis may be performed on the expression vector. For example, *ccdB* counterselection combined with Red/ET recombination in *E. coli* could seamlessly introduce mutations in plasmids [115, 139]. Zinc-finger nucleases (ZFN) [140], the newly-developed transcription activator-like effector nucleases (TALENs) [141] and the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) systems [142] also have potential in editing secondary metabolites biosynthetic gene clusters. After a series of genetic engineering, the expression vector could be introduced into heterologous hosts by transformation, transduction or conjugation. The transformants are then cultivated in appropriate media and the produced compounds could be detected by liquid chromatography–mass spectrometry (LC-MS). The production of target compounds could be validated by comparing the retention time, MS and MS/MS fragments with reference compounds. In order to verify the chemical structure, target compounds may need to be purified and measured by nuclear magnetic resonance (NMR). In case of low production in heterologous hosts, the promoter probably has to be changed to ensure effective transcription of the biosynthetic gene cluster, and optimization of metabolic pathways in heterologous hosts may be required to drive appropriate substrates to secondary metabolites biosynthesis. Employment of other heterologous hosts may also be taken into account. Conversely, if the

heterologous production is successful, the production titer could be improved further by optimizing fermentation conditions.
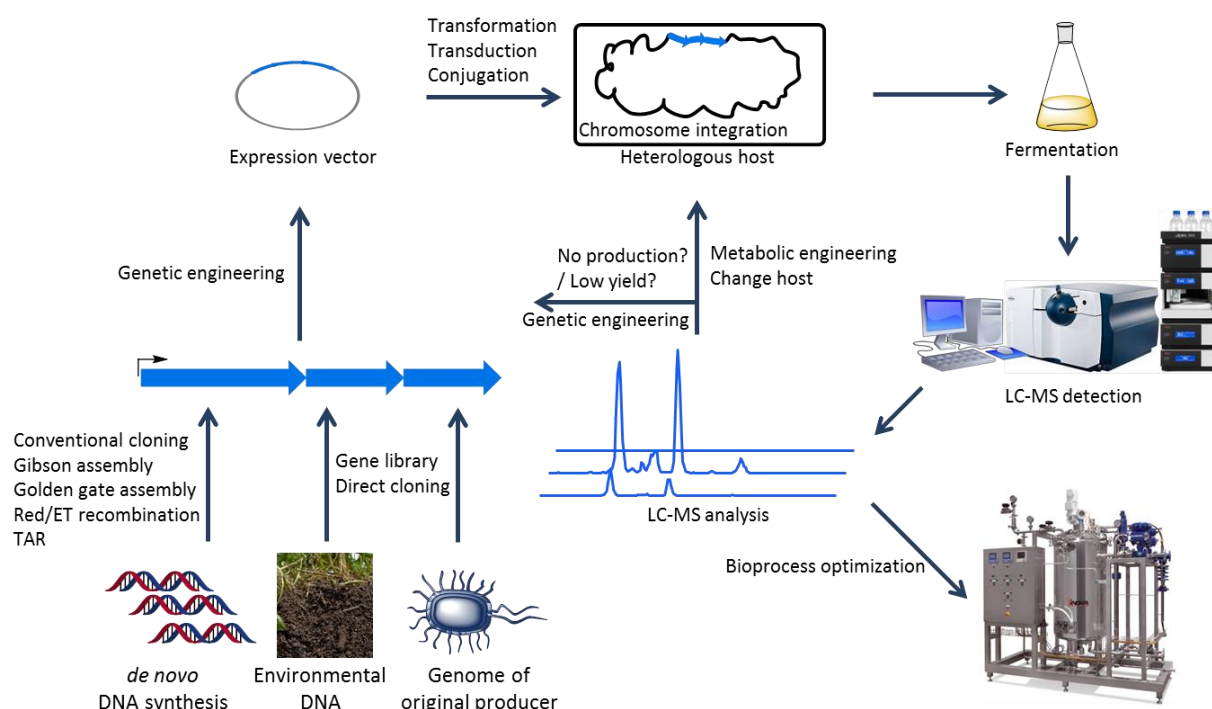
Fig. 1.7 Scheme of heterologous expression.

Driven by the progress in large DNA fragment cloning and DNA synthesis, several PKs and NRPs biosynthetic gene clusters have been successfully cloned and expressed in heterologous hosts. Take myxobacterial natural products for example, epothilone (anticancer), myxothiazole (respiration inhibitor), myxochromide S (sufactant), pretubulysin (anticancer) and disorazol (anticancer) were successfully produced in heterologous host *Myxococcous xanthus* [143–148]. Microbes like *Streptomyces*, *E. coli*, *Bacillus subtilis*, *Pseudomonas*, *Mycobacterium smegmatis*, *Burkholderia*, *Aspergillus nidulans* and *Saccharomyces cerevisiae* etc. also have been used as heterologous host to produce secondary metabolites [130, 149–161]. Moreover, Sf9 insect cells and even plants are also available to serve as heterologous host [151]. Although they have been excavated for nearly a century, bacteria still have great potential in

antibiotics discovery. To date only a small fraction of bacteria could be cultivated in laboratory conditions [162]. Due to the development of metagenome sequencing [163–165] and single-cell genome sequencing [166], the genetic information of novel antibiotics biosynthetic gene clusters in unculturable microorganisms became accessible. Introducing the biosynthetic gene cluster of the target compound into heterologous host could be a versatile method to explore this hidden world.

1.4 Myxobacteria and natural products

Myxobacteria are Gram-negative bacteria and most of them are soil dwelling microbes [167]. They belong to δ-proteobacteria in the order *Myxococcales* and could behave in both unicellular and multicellular forms [168]. One of their characteristics is gliding and swarming on surfaces of matrix (Fig. 1.8). Under starvation vegetative cells differentiate into myxospores and form fruiting bodies. They are also renowned for predatory of other microbes by producing exoenzymes and bioactive small molecules [168–172]. Although the earliest discovery of myxobacteria was in 1892 [173], extensive study on this kind of bacteria started from 1940s. Usually myxobacteria are slow growing in laboratory conditions, and most strains could not grow in homogenous suspensions, so in many cases it requires a long period to characterize natural products from myxobacteria [167]. Over the last decades myxobacteria have grown up to be an important source of secondary metabolites besides of actinomycetes, bacillus, pseudomonads and fungi. To date more than 100 core structures and around 600 derivatives have been identified from myxobacteria [174]. The bioactivities of these compounds are diverse from anti-infective to cytotoxic, and their modes of actions (MOA) are distinctive [175–177] (Fig. 1.9).

1. Introduction

Since the genome of *Myxococcus xanthus* DK1622 been sequenced in 2006 [178], around 25 myxobacteria genomes have been sequenced. Their genome sequences are available in the NCBI Genome database. Most of their genome sizes are greater than 5 Mb. The genome size of *Sorangium cellulosum* So ce56 even reached up to 13 Mb. The GC content of the myxobacteria genomes is generally very high (63% ~ 75%). To date, around forty biosynthetic gene clusters from myxobacteria have been published, and around half of them are PKS/NRPS hybrids (Table 1.2). Deeper investigations on genome, proteome and metabolome of myxobacteria uncovered their great potential in novel structure discovery and biosynthetic pathway expression [164, 179–182].

| *Myxococcus xanthus*
swarming colony | *Sorangium sp.*
swarming colony | *Stigmatella aurantiaca*
fruiting bodies | *Chondromyces apiculatus*
fruiting bodies |
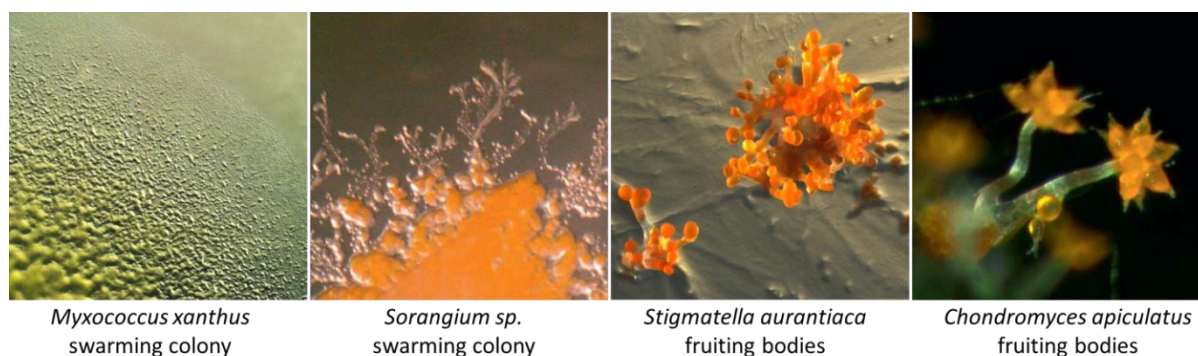
Fig. 1.8 Myxobacteria colonies and fruiting bodies.

Table 1.2 Biosynthetic gene clusters identified from myxobacteria*

Compound	Producer	Biosynthetic Type
Saframycin	*Myxococcus xanthus* DSM504/15	NRPS
Soraphen	*Sorangium cellulosum* So ce26	PKS
Myxovirescin	*Myxococcus xanthus* DK1622	PKS/NRPS
Myxothiazol	*Stigmatella aurantiaca* DW4/3-1	PKS/NRPS
Epothilone	*Sorangium cellulosum* So ce90	PKS/NRPS
Myxochelin	*Stigmatella aurantiaca* Sg a15	NRPS
Myxalamid	*Stigmatella aurantiaca* Sg a15	PKS/NRPS
Stigmatellin	*Stigmatella aurantiaca* Sg a15	PKS
Melithiazol	*Melittangium lichenicola* Me l46	PKS/NRPS

Tubulysin	*Angiococcus disciformis* An d48	PKS/NRPS
Disorazol	*Sorangium cellulosum* So ce12	PKS/NRPS
Myxochromide S	*Stigmatella aurantiaca* DW4/3-1	PKS/NRPS
Cystothiazole	*Cystobacter fuscus* AJ-13278	PKS/NRPS
Chivosazol	*Sorangium cellulosum* So ce56	PKS/NRPS
Myxochromide A	*Myxococcus xanthus* DK1622	PKS/NRPS
Chondramide	*Chondromyces crocatus* Cm c5	PKS/NRPS
Ambruticin	*Sorangium cellulosum* So ce10	PKS/NRPS
Jerangolid	*Sorangium cellulosum* So ce307	PKS
Aurachin	*Stigmatella aurantiaca* Sg a15	PKS
Spirangien	*Sorangium cellulosum* So ce90	PKS
Etnangien	*Sorangium cellulosum* So ce56	PKS
DKxanthene	*Myxococcus xanthus* DK1622 and *M. xanthus* DK1050	PKS/NRPS
Aurafuron	*Stigmatella aurantiaca* DW4/3-1	PKS
Ajudazol	*Chondromyces crocatus* Cm c5	PKS/NRPS
Chondrochloren	*Chondromyces crocatus* Cm c5	PKS/NRPS
Thuggacin	*Sorangium cellulosum* So ce895 and *Chondromyces crocatus* Cm c5	PKS/NRPS
Leupyrrins	*Sorangium cellulosum* So ce690	PKS/NRPS
Rhizopodin	*Stigmatella aurantiaca* Sg a15	PKS/NRPS
Myxopyronin	*Myxococcus fulvus* Mx f50	PKS
Corallopyronin	*Corallococcus coralloides* B035	PKS/NRPS
Sorangicin	*S. cellulosum* So ce12	PKS
Microsclerodermins	*Sorangium* and *Jahnella* species	PKS/NRPS
Gephyronic acid	*Cystobacter violaceus* Cb vi76	PKS
Phenylnannolone A	*Nannocystis pusilla* B150	PKS
Crocacin	*Chondromyces crocatus* Cm c5	PKS/NRPS
Cystomanamide	*Cystobacter fuscus* MCy9118	PKS/NRPS
Dawenol	*Stigmatella aurantiaca* DW4/3-1	PKS
Macyranone	*Cystobacter fuscus* MCy9118	PKS/NRPS
Bengamide	*Myxococcus virescens* ST200611	PKS/NRPS
Chlorotonil	*Sorangium cellulosum* So ce1525	PKS
Haliangicin	*Haliangium ochraceum* SMP-2	PKS

*, the table is summarized from [98, 180, 181, 183–198].
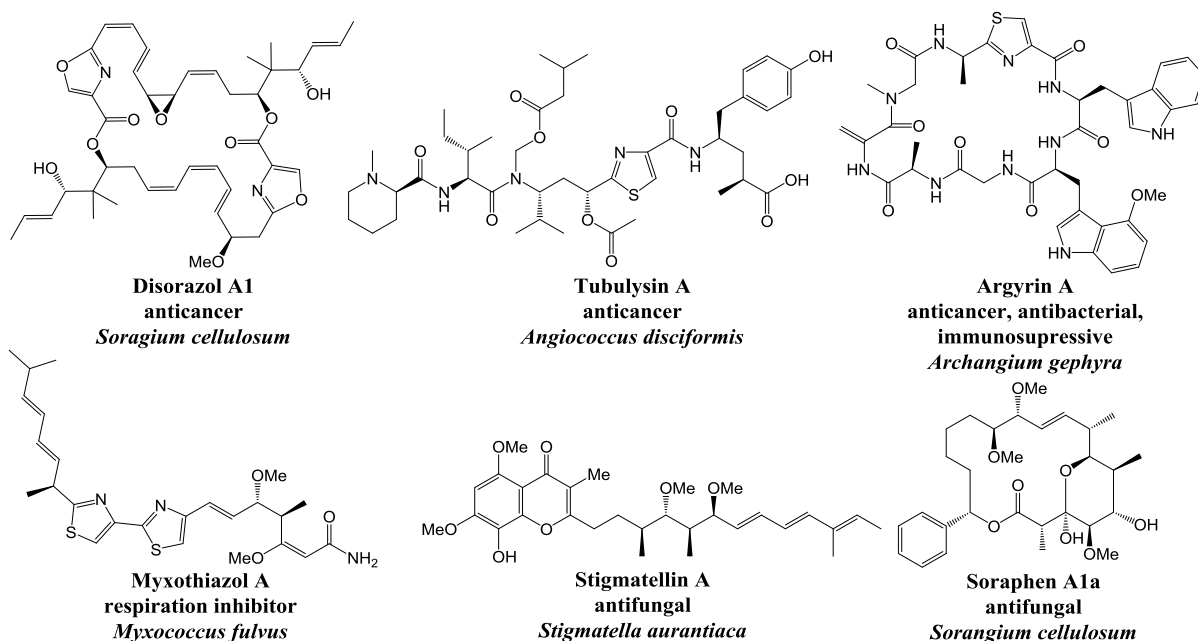
Fig. 1.9 Selected bioactive compounds from myxobacteria.

Disorazol A1
anticancer
Soragium cellulosum

Tubulysin A
anticancer
Angiococcus disciformis

Argyrin A
anticancer, antibacterial,
immunosupressive
Archangium gephyra

Myxothiazol A
respiration inhibitor
Myxococcus fulvus

Stigmatellin A
antifungal
Stigmatella aurantiaca

Soraphen A1a
antifungal
Sorangium cellulosum

1.5 Outline of this dissertation

The topic of this thesis dealt with biosynthesis and heterologous production of vioprolides from the myxobacterium *Cystobacter violaceus* Cb vi35 (Fig. 1.10). Vioprolides are cyclic peptolides with remarkable anticancer and antifungal activities [199] (Fig. 1.11). An unprecedented moiety *trans*-(*2S, 4R*)-4-methylazetidine-carboxylic acid (Maz) is harbored by vioprolide A and C, while L-glyceric acid is also uncommon in nature. Vioprolides A, B and C are found highly cytotoxic to mammalian cells but with low anti-fungal activity. Conversely, vioprolide D shows high activity to a broad spectrum of fungi and yeasts but is less toxic against mammalian cells. Consequently, they have great potential to be developed as anti-fungal drugs and anti-cancer agents. In addition, vioprolides are also possible to be used to treat various diseases and disorders by enhancing the level of interferon α and β (US Patent: US 2010/0028298 A1 and US20110173707). Due to the low amount of vioprolides produced by the native producer, research on their mode of action has not met with success. Although

several structural fragments of vioprolides such as *E*-Dhb-Thiazoline-Pro-Leu and azetidinyl-thiazoline fragments could be synthesized chemically [200, 201], total synthesis of vioprolides is not established yet.

C. Violaceus Cb vi35
vegetative cells

C. Violaceus Cb vi35
swarming colony

C. Violaceus Cb vi35
fruiting bodies

C. Violaceus Cb vi35
fruiting bodies
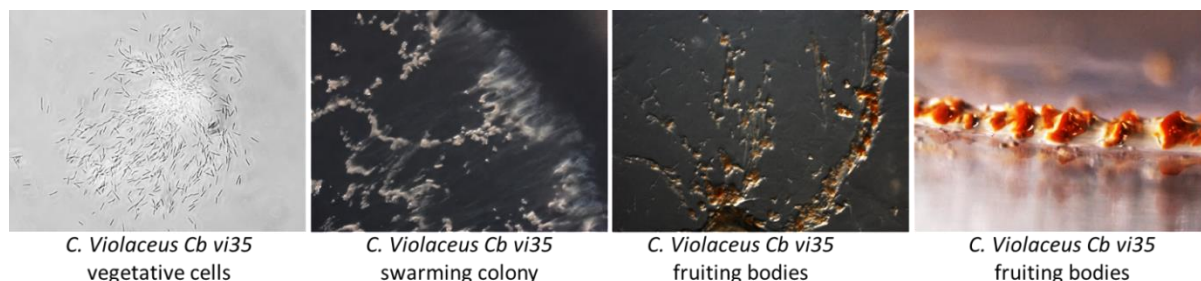
Fig. 1.10 Different life stages of *C. violaceus* Cb vi35.

Vioprolide A   **Vioprolide B**   **Vioprolide C**   **Vioprolide D**
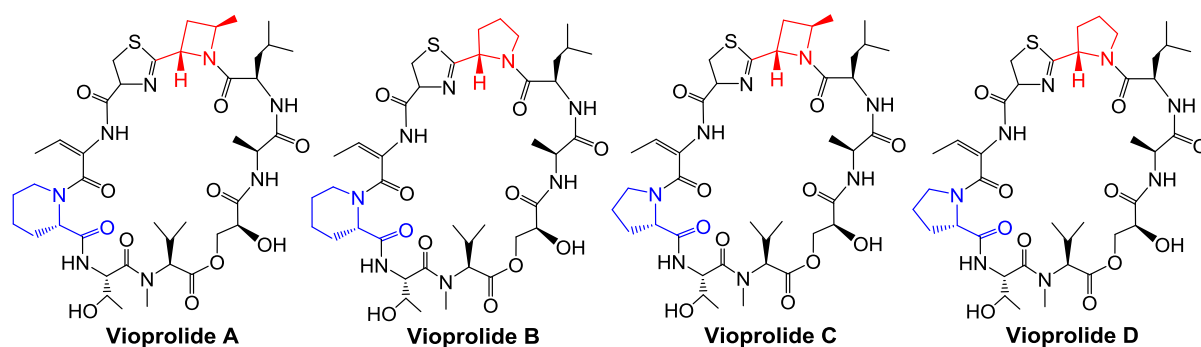
Fig. 1.11 Chemical structures of vioprolides.

Since the isolation of vioprolides from the myxobacterium *C. violaceus* Cb vi35, little is known about their biosynthetic mechanism. In this dissertation, the in-depth sequence analysis and retro-biosynthesis prediction on the Cb vi35 genome gave insights to the vioprolide biosynthetic gene cluster. In order to identify the previoprolides, we performed isotope labeling in *C. violaceus* Cb vi35. For the purpose of confirming the vioprolide biosynthetic gene cluster, we tried to clone the *vio* gene cluster from Cb vi35 genome and express it in heterologous hosts. Constitutive promoters were applied to drive the *vio* gene cluster in heterologous systems to increase the production yield, and site-directed mutagenesis was carried out to generate vioprolide analogues. Meanwhile, we tried to isolate vioprolide derivatives and previoprolides from

heterologous system and verify their structures by NMR. Furthermore, *in vitro* biochemical characterizations were performed to investigate the mechanism of assembly initiation and the unusual glycerate esterification process. The results are described and discussed in **Section 2** and **Section 3**.

1.6 References

1   Patridge, E., Gareiss, P., Kinch, M.S., Hoyer, D. (2016) An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discovery Today*, **21** (2), 204–207.

2   Newman, D.J. and Cragg, G.M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal of Natural Products*, **75** (3), 311–335.

3   Newman, D.J. and Cragg, G.M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*, **79** (3), 629–661.

4   Newman, D.J. and Cragg, G.M. (2009) Natural product scaffolds as leads to drugs. *Future Medicinal Chemistry*, **1** (8), 1415–1427.

5   Fleming, A. (1929) On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. influenzæ*. *British Journal of Experimental Pathology*, **10** (3), 226–236.

6   Aminov, R.I. (2010) A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in Microbiology*, **1**, 134.

7   Davies, J. and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews : MMBR*, **74** (3), 417–433.

8   World Health Organization Antimicrobial Resistance: Global report on surveillance (2014).

9   World Health Organization Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance and response.

10 Fair, R.J. and Tor, Y. (2014) Antibiotics and bacterial resistance in the 21st century. *Perspectives in Medicinal Chemistry*, **6**, 25–64.

11 Hede, K. (2014) Antibiotic resistance: An infectious arms race. *Nature*, **509** (7498), S2-S3.

12 Morgan, J. (2005) Global trends in candidemia: review of reports from 1995-2005. *Current Infectious Disease Reports*, **7** (6), 429–439.

13 Wisplinghoff, H., Bischoff, T., Tallent, S.M., Seifert, H., Wenzel, R.P., Edmond, M.B. (2004) Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, **39** (3), 309–317.

14 Shlaes, D.M. (2010) *Antibiotics: The perfect storm*, Springer, Dordrecht, New York.

15 Cochrane, S.A. and Vederas, J.C. (2016) Lipopeptides from *Bacillus* and *Paenibacillus spp.*: A Gold Mine of Antibiotic Candidates. *Medicinal Research Reviews*, **36** (1), 4–31.

16 Meka, V.G. and Gold, H.S. (2004) Antimicrobial resistance to linezolid. *Clinical Infectious Diseases: an official publication of the Infectious Diseases Society of America*, **39** (7), 1010–1015.

17 Munoz-Price, L.S., Lolans, K., Quinn, J.P. (2005) Emergence of resistance to daptomycin during treatment of vancomycin-resistant *Enterococcus faecalis* infection. *Clinical Infectious Diseases: an official publication of the Infectious Diseases Society of America*, **41** (4), 565–566.

18 Vikram, H.R., Havill, N.L., Koeth, L.M., Boyce, J.M. (2005) Clinical progression of methicillin-resistant *Staphylococcus aureus* vertebral osteomyelitis associated with reduced susceptibility to daptomycin. *Journal of Clinical Microbiology*, **43** (10), 5384–5387.

19 Stewart, B.W. and Wild, C. (eds) (2014) *World cancer report 2014*, International Agency for Research on Cancer; WHO Press World Health Organization, Lyon France, Geneva Switzerland.

20 Side Effects in cancer treatment. http://www.cancer.gov/about-cancer/treatment/side-effects.

21 Leget, G.A. and Czuczman, M.S. (1998) Use of rituximab, the new FDA-approved antibody. *Current Opinion in Oncology*, **10** (6), 548–551.

22 Gerber, D.E. (2008) Targeted therapies: a new generation of cancer treatments. *American Family Physician*, **77** (3), 311–319.

23 Abramson, R. 2016. Overview of Targeted Therapies for Cancer. *My Cancer Genome.*

https://www.mycancergenome.org/content/molecular-medicine/overview-of-targeted-therapies-for-cancer/

24 Wu, J., Joseph, S.O., Muggia, F.M. (2012) Targeted therapy: its status and promise in selected solid tumors part I: areas of major impact. *Oncology (Williston Park, N.Y.)*, **26** (10), 936–943.

25 Joseph, S.O., Wu, J., Muggia, F.M. (2012) Targeted therapy: its status and promise in selected solid tumors. Part II: Impact on selected tumor subsets, and areas of evolving integration. *Oncology (Williston Park, N.Y.)*, **26** (11), 1021.

26 Hoelder, S., Clarke, P.A., Workman, P. (2012) Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Molecular Oncology*, **6** (2), 155–176.

27 Druker, B.J., Talpaz, M., Resta, D.J., Peng, B., Buchdunger, E., Ford, J.M., Lydon, N.B., Kantarjian, H., Capdeville, R., Ohno-Jones, S., Sawyers, C.L. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *The New England Journal of Medicine*, **344** (14), 1031–1037.

28 Musumeci, F., Schenone, S., Grossi, G., Brullo, C., Sanna, M. (2015) Analogs, formulations and derivatives of imatinib: a patent review. *Expert Opinion on Therapeutic Patents*, **25** (12), 1411–1421.

29 Trivedi, M., Budihardjo, I., Loureiro, K., Reid, T.R., Ma, J.D. (2008) Epothilones: a novel class of microtubule-stabilizing drugs for the treatment of cancer. *Future Oncology (London, England)*, **4** (4), 483–500.

30 Kim, C.H., Axup, J.Y., Lawson, B.R., Yun, H., Tardif, V., Choi, S.H., Zhou, Q., Dubrovska, A., Biroc, S.L., Marsden, R., Pinstaff, J., Smider, V.V., Schultz, P.G. (2013) Bispecific small molecule-antibody conjugate targeting prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (44), 17796–17801.

31 Perez, H.L., Cardarelli, P.M., Deshpande, S., Gangwar, S., Schroeder, G.M., Vite, G.D., Borzilleri, R.M. (2014) Antibody-drug conjugates: current status and future directions. *Drug Discovery Today*, **19** (7), 869–881.

32 Thomas, A., Teicher, B.A., Hassan, R. (2016) Antibody-drug conjugates for cancer therapy. *The Lancet. Oncology*, **17** (6), 254-262.

33 Zolot, R.S., Basu, S., Million, R.P. (2013) Antibody-drug conjugates. *Nature reviews. Drug Discovery*, **12** (4), 259–260.

34 Demain, A.L. and Fang, A. (2000) The natural functions of secondary metabolites. *Advances in Biochemical Engineering/Biotechnology*, **69**, 1–39.

35 Walsh, C.T. and Fischbach, M.A. (2010) Natural products version 2.0: connecting genes to molecules. *Journal of the American Chemical Society*, **132** (8), 2469–2493.

36 Fischbach, M.A. and Walsh, C.T. (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews*, **106** (8), 3468–3496.

37 Cane, D.E. (1998) Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science*, **282** (5386), 63–68.

38 Schwarzer, D. and Marahiel, M.A. (2001) Multimodular biocatalysts for natural product assembly. *Naturwissenschaften*, **88** (3), 93–101.

39 KIRBY, R., Wright, L.F., Hopwood, D.A. (1975) Plasmid-determined antibiotic synthesis and resistance in *Streptomyces coelicolor*. *Nature*, **254** (5497), 265–267.

40 Wright, L.F. and Hopwood, D.A. (1976) Identification of the antibiotic determined by the SCPI plasmid of *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, **95**, 96–106.

41 Wright, L.F. and Hopwood, D.A. (1976) Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolar* A3(2). *Journal of General Microbiology*, **96** (2), 289–297.

42 Malpartida, F. and Hopwood, D.A. (1984) Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature*, **309** (5967), 462–464.

43 Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A., Floss, H.G., Omura, S. (1985) Production of 'hybrid' antibiotics by genetic engineering. *Nature*, **314** (6012), 642–644.

44 Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F., Hutchinson, C.R. (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *The EMBO journal*, **8** (9), 2727–2736.

45 Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.M., Hopwood, D.A. (1989) Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. *The EMBO journal*, **8** (9), 2717–2725.

46 Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., Leadlay, P.F. (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature*, **348** (6297), 176–178.

47 Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science (New York, N.Y.)*, **252** (5006), 675–679.

48 Du, L. and Lou, L. (2010) PKS and NRPS release mechanisms. *Natural Product Reports*, **27** (2), 255–278.

49 Sundaram, S. and Hertweck, C. (2016) On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems. *Current Opinion in Chemical Biology*, **31**, 82–94.

50 Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angewandte Chemie (International ed. in English)*, **48** (26), 4688–4716.

51 Gevers, W., Kleinkauf, H., Lipmann, F. (1968) The activation of amino acids for biosynthesis of gramicidin S. *Proceedings of the National Academy of Sciences of the United States of America*, **60** (1), 269–276.

52 Gevers, W., Kleinkauf, H., Lipmann, F. (1969) Peptidyl transfers in gramicidin S bisoynthesis from enzyme-bound thioester intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, **63** (4), 1335–1342.

53 Kleinkauf, H., Roskoski, R., JR, Lipmann, F. (1971) Pantetheine-linked peptide intermediates in gramicidin S and tyrocidine biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **68** (9), 2069–2072.

54 Lee, S.G. and Lipmann, F. (1974) Isolation of a peptidyl-pantetheine-protein from tyrocidine-synthesizing polyenzymes. *Proceedings of the National Academy of Sciences of the United States of America*, **71** (3), 607–611.

55 Pollack, J.R., Ames, B.N., Neilands, J.B. (1970) Iron transport in *Salmonella typhimurium*: mutants blocked in the biosynthesis of enterobactin. *Journal of Bacteriology*, **104** (2), 635–639.

56 Luke, R.K. and Gibson, F. (1971) Location of three genes concerned with the conversion of 2, 3-dihydroxybenzoate into enterochelin in *Escherichia coli* K-12. *Journal of Bacteriology*, **107** (2), 557–562.

57 Young, I.G., Langman, L., Luke, R.K., Gibson, F. (1971) Biosynthesis of the iron-transport compound enterochelin: mutants of *Escherichia coli* unable to synthesize 2, 3-dihydroxybenzoate. *Journal of Bacteriology*, **106** (1), 51–57.

58 Marahiel, M.A., Krause, M., Skarpeid, H.J. (1985) Cloning of the tyrocidine synthetase 1 gene from *Bacillus brevis* and its expression in *Escherichia coli*. *Molecular & General Genetics: MGG*, **201** (2), 231–236.

59 Krause, M. and Marahiel, M.A. (1988) Organization of the biosynthesis genes for the peptide antibiotic gramicidin S. *Journal of Bacteriology*, **170** (10), 4669–4674.

60 Nakano, M.M., Marahiel, M.A., Zuber, P. (1988) Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *Journal of Bacteriology*, **170** (12), 5662–5668.

61 Kratzschmar, J., Krause, M., Marahiel, M.A. (1989) Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. *Journal of Bacteriology*, **171** (10), 5422–5429.

62 Mittenhuber, G., Weckermann, R., Marahiel, M.A. (1989) Gene cluster containing the genes for tyrocidine synthetases 1 and 2 from *Bacillus brevis*: evidence for an operon. *Journal of Bacteriology*, **171** (9), 4881–4887.

63 Rusnak, F., Faraci, W.S., Walsh, C.T. (1989) Subcloning, expression, and purification of the enterobactin biosynthetic enzyme 2, 3-dihydroxybenzoate-AMP ligase: Demonstration of enzyme-bound (2,3-dihydroxybenzoyl)adenylate product. *Biochemistry*, **28** (17), 6827–6835.

64 Liu, J., Quinn, N., Berchtold, G.A., Walsh, C.T. (1990) Overexpression, purification and characterization of isochorismate synthase (EntC), the first enzyme involved in the biosynthesis of enterobactin from chorismate. *Biochemistry*, **29** (6), 1417–1425.

65 Rusnak, F., Liu, J., Quinn, N., Berchtold, G.A., Walsh, C.T. (1990) Subcloning of the enterobactin biosynthetic gene *entB*: Expression, purification, characterization and substrate specificity of isochorismatase. *Biochemistry*, **29** (6), 1425–1435.

66 Walsh, C.T., Liu, J., Rusnak, F., Sakaitani, M. (1990) Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chemical Review*, **90** (7), 1105–1129.

67 Rusnak, F., Sakaitani, M., Drueckhammer, D., Reichert, J., Walsh, C.T. (1991) Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *entF* gene, expression and purification of EntF, and analysis of covalent phosphopantetheine. *Biochemistry*, **30** (11), 2916–2927.

68 Reichert, J., Sakaitani, M., Walsh, C.T. (1992) Characterization of EntF as a serine-activating enzyme. *Protein Science: a publication of the Protein Society*, **1** (4), 549–556.

69 Lambalot, R.H. and Walsh, C.T. (1995) Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *The Journal of Biological Chemistry*, **270** (42), 24658–24661.

70 Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., Walsh, C.T. (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & Biology*, **3** (11), 923–936.

71 Gehring, A.M., Bradley, K.A., Walsh, C.T. (1997) Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that

is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry*, **36** (28), 8495–8503.

72 Gehring, A.M., Mori, I., Walsh, C.T. (1998) Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry*, **37** (8), 2648–2659.

73 Shaw-Reid, C.A., Kelleher, N.L., Losey, H.C., Gehring, A.M., Berg, C., Walsh, C.T. (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: The thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chemistry & Biology*, **6** (6), 385–400.

74 Walsh, C.T. (2016) Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Natural Product Reports*, **33** (2), 127–135.

75 Horsman, M.E., Hari, T.P.A., Boddy, C.N. (2016) Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Natural Product Reports*, **33** (2), 183–202.

76 Rausch, C., Hoof, I., Weber, T., Wohlleben, W., Huson, D.H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, **7**, 78.

77 Balibar, C.J., Vaillancourt, F.H., Walsh, C.T. (2005) Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chemistry & Biology*, **12** (11), 1189–1200.

78 Reimmann, C., Patel, H.M., Serino, L., Barone, M., Walsh, C.T., Haas, D. (2001) Essential PchG-dependent reduction in pyochelin biosynthesis of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, **183** (3), 813–820.

79 Hansen, D.B., Bumpus, S.B., Aron, Z.D., Kelleher, N.L., Walsh, C.T. (2007) The loading module of mycosubtilin: an adenylation domain with fatty acid selectivity. *Journal of the American Chemical Society*, **129** (20), 6366–6367.

80 Boettger, D. and Hertweck, C. (2013) Molecular diversity sculpted by fungal PKS-NRPS hybrids. *ChemBioChem : a European journal of chemical biology*, **14** (1), 28–42.

81 Shen, B., Chen, M., Cheng, Y., Du, L., Edwards, D.J., George, N.P., Huang, Y., Oh, T., Sanchez, C., Tang, G., Wendt-Pienkowski, E., Yi, F. (2005) Prerequisites for Combinatorial Biosynthesis: Evolution of Hybrid NRPS/PKS Gene Clusters, in *Biocombinatorial Approaches for Drug Finding* (eds W. Wohlleben, T. Spellig, B. Müller-Tiemann), Springer, Berlin, New York, pp. 107–126.

82 Du, L., Sanchez, C., Shen, B. (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic Engineering*, **3** (1), 78–95.

83 Hamley, I.W. (2015) Lipopeptides: from self-assembly to bioactivity. *Chemical Communications (Cambridge, England)*, **51** (41), 8574–8583.

84 Ongena, M. and Jacques, P. (2008) Bacillus lipopeptides: versatile weapons for plant disease biocontrol. *Trends Microbiology*, **16** (3), 115–125.

85 McPhail, K.L., Correa, J., Linington, R.G., Gonzalez, J., Ortega-Barria, E., Capson, T.L., Gerwick, W.H. (2007) Antimalarial linear lipopeptides from a Panamanian strain of the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, **70** (6), 984–988.

86 Raaijmakers, J.M., Bruijn, I. de, de Kock, Maarten J D (2006) Cyclic lipopeptide production by plant-associated *Pseudomonas spp.*: diversity, activity, biosynthesis, and regulation. *Molecular Plant-Microbe Interactions : MPMI*, **19** (7), 699–710.

87 Kugler, J.H., Le Roes-Hill, M., Syldatk, C., Hausmann, R. (2015) Surfactants tailored by the class Actinobacteria. *Frontiers in Microbiology*, **6**, 212.

88 Hiraoka, H., Ano, T., Shoda, M. (1992) Molecular cloning of a gene responsible for the biosynthesis of the lipopeptide antibiotics iturin and surfactin. *Journal of Fermentation and Bioengineering*, **74** (5), 323–326.

89 Huang, C.C., Ano, T., Shoda, M. (1993) Nucleotide sequence and characteristics of the gene, *lpa-14*, responsible for biosynthesis of the lipopeptide antibiotics iturin A and surfactin from *Bacillus subtilis* RB14. *Journal of Fermentation and Bioengineering*, **76** (6), 445–450.

90 Cosmina, P., Rodriguez, F., Deferra, F., Grandi, G., Perego, M., Venema, G., Vansinderen, D. (1993) Sequence and analysis of the genetic-locus responsible for surfactin synthesis in *Bacillus subtilis*. *Molecular Microbiology*, **8** (5), 821–831.

91 Menkhaus, M., Ullrich, C., Kluge, B., Vater, J., Vollenbroich, D., Kamp, R.M. (1993) Structural and functional organization of the surfactin synthetase multienzyme system. *Journal of Biological Chemistry*, **268** (11), 7678–7684.

92 Peypoux, F., Bonmatin, J.M., Wallach, J. (1999) Recent trends in the biochemistry of surfactin. *Applied Microbiology and Biotechnology*, **51** (5), 553–563.

93 Baltz, R.H., Miao, V., Wrigley, S.K. (2005) Natural products to drugs: daptomycin and related lipopeptide antibiotics. *Natural Product Reports*, **22** (6), 717–741.

94 Miao, V., Coeffet-LeGal, M.F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C.J., Wrigley, S.K., Baltz, R.H. (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: Cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology*, **151** (Pt 5), 1507–1523.

95 Steenbergen, J.N., Alder, J., Thorne, G.M., Tally, F.P. (2005) Daptomycin: a lipopeptide antibiotic for the treatment of serious Gram-positive infections. *J Antimicrob Chemother*, **55** (3), 283–288.

96 Robbel, L. and Marahiel, M.A. (2010) Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *Journal of Biological Chemistry*, **285** (36), 27501–27508.

97 Fu, C., Keller, L., Bauer, A., Brönstrup, M., Froidbise, A., Hammann, P., Herrmann, J., Mondesert, G., Kurz, M., Schiell, M., Schummer, D., Toti, L., Wink, J., Müller, R. (2015) Biosynthetic Studies of telomycin reveal new lipopeptides with enhanced activity. *Journal of the American Chemical Society*, **137** (24), 7692–7705.

98 Etzbach, L., Plaza, A., Garcia, R., Baumann, S., Müller, R. (2014) Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Organic Letters*, **16** (9), 2414–2417.

99 Wenzel, S.C., Kunze, B., Höfle, G., Silakowski, B., Scharfe, M., Blocker, H., Müller, R. (2005) Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *ChemBioChem : a European journal of chemical biology*, **6** (2), 375–385.

100 Reimer, D., Pos, K.M., Thines, M., Grün, P., Bode, H.B. (2011) A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nature Chemical Biology*, **7** (12), 888–890.

101 Brotherton, C.A. and Balskus, E.P. (2013) A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *Journal of the American Chemical Society*, **135** (9), 3359–3362.

102 Kevany, B.M., Rasko, D.A., Thomas, M.G. (2009) Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. *Applied and Environmental Microbiology*, **75** (4), 1144–1155.

103   Xu, Y., Kersten, R.D., Nam, S.-J., Lu, L., Al-Suwailem, A.M., Zheng, H., Fenical, W., Dorrestein, P.C., Moore, B.S., Qian, P.-Y. (2012) Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *Journal of the American Chemical Society*, **134** (20), 8625–8632.

104   Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A., Smith, C.P., Micklefield, J. (2002) Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chemistry & Biology*, **9** (11), 1175–1187.

105   Wessels, P., Dohren, H. von, Kleinkauf, H. (1996) Biosynthesis of acylpeptidolactones of the daptomycin type. A comparative analysis of peptide synthetases forming A21978C and A54145. *European Journal of Biochemistry / FEBS*, **242** (3), 665–673.

106   Kurth, C., Schieferdecker, S., Athanasopoulou, K., Seccareccia, I., Nett, M. (2016) Variochelins, Lipopeptide Siderophores from Variovorax boronicumulans Discovered by Genome Mining. *Journal of Natural Products*, **79** (4), 865–872.

107   Luo, C., Liu, X., Zhou, H., Wang, X., Chen, Z. (2015) Nonribosomal peptide synthase gene clusters for lipopeptide biosynthesis in *Bacillus subtilis* 916 and their phenotypic functions. *Applied and Environmental Microbiology*, **81** (1), 422–431.

108   Mares, J., Hajek, J., Urajova, P., Kopecky, J., Hrouzek, P. (2014) A hybrid non-ribosomal peptide/polyketide synthetase containing fatty-acyl ligase (FAAL) synthesizes the beta-amino fatty acid lipopeptides puwainaphycins in the Cyanobacterium *Cylindrospermum alatosporum*. *PLoS ONE*, **9** (11), e111904.

109   Winn, M., Fyans, J.K., Zhuo, Y., Micklefield, J. (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Natural Product Reports*, **33** (2), 317–347.

110   McDaniel, R., Welch, M., Hutchinson, C.R. (2005) Genetic approaches to polyketide antibiotics. 1. *Chemical Reviews*, **105** (2), 543–558.

111   Bilyk, O. and Luzhetskyy, A. (2016) Metabolic engineering of natural product biosynthesis in actinobacteria. *Current Opinion in Biotechnology*, **42**, 98–107.

112   Kealey, J.T. (2003) Creating polyketide diversity through genetic engineering. *Frontiers in Bioscience*, **8** (3), c1-13.

113   Miao, V., Coeffet-Le Gal, M.-F., Nguyen, K., Brian, P., Penn, J., Whiting, A., Steele, J., Kau, D., Martin, S., Ford, R., Gibson, T., Bouchard, M., Wrigley, S.K., Baltz, R.H. (2006) Genetic engineering in *Streptomyces*

roseosporus to produce hybrid lipopeptide antibiotics. *Chemistry & Biology*, **13** (3), 269–276.

114   Doekel, S., Coeffet-Le Gal, M.-F., Gu, J.-Q., Chu, M., Baltz, R.H., Brian, P. (2008) Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiology (Reading, England)*, **154** (Pt 9), 2872–2880.

115   Bian, X., Plaza, A., Yan, F., Zhang, Y., Müller, R. (2015) Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives *in vivo*. *Biotechnology and Bioengineering*, **112** (7), 1343–1353.

116   Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D., Hilvert, D. (2014) Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angewandte Chemie (International ed. in English)*, **53** (38), 10105–10108.

117   Crusemann, M., Kohlhaas, C., Piel, J. (2013) Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chemical Science*, **4** (3), 1041–1045.

118   Calcott, M.J. and Ackerley, D.F. (2014) Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnology Letters*, **36** (12), 2407–2416.

119   Williams, G.J. (2013) Engineering polyketide synthases and nonribosomal peptide synthetases. *Current Opinion in Structural Biology*, **23** (4), 603–612.

120   Hans, M., Hornung, A., Dziarnowski, A., Cane, D.E., Khosla, C. (2003) Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules. *Journal of the American Chemical Society*, **125** (18), 5366–5374.

121   Kato, Y., Bai, L., Xue, Q., Revill, W.P., Yu, T.-W., Floss, H.G. (2002) Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. *Journal of the American Chemical Society*, **124** (19), 5268–5269.

122   Stassi, D.L., Kakavas, S.J., Reynolds, K.A., Gunawardana, G., Swanson, S., Zeidner, D., Jackson, M., Liu, H., Buko, A., Katz, L. (1998) Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **95** (13), 7305–7309.

123   Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R., McDaniel, R. (2001) Alteration of the substrate specificity of a modular

polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry*, **40** (51), 15464–15470.

124   Dunn, B.J. and Khosla, C. (2013) Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *Journal of the Royal Society, Interface / the Royal Society*, **10** (85), 20130297.

125   Brautaset, T., Sletta, H., Nedal, A., Borgos, S.E.F., Degnes, K.F., Bakke, I., Volokhan, O., Sekurova, O.N., Treshalin, I.D., Mirchink, E.P., Dikiy, A., Ellingsen, T.E., Zotchev, S.B. (2008) Improved antifungal polyene macrolides via engineering of the nystatin biosynthetic genes in *Streptomyces noursei. Chemistry & Biology*, **15** (11), 1198–1206.

126   Kim, H.U., Charusanti, P., Lee, S.Y., Weber, T. (2016) Metabolic engineering with systems biology tools to optimize production of prokaryotic secondary metabolites. *Natural Product Reports*, 33 (8), 933-941.

127   Zarins-Tutt, J.S., Barberi, T.T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D.J., Goss, R.J.M. (2016) Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Natural Product Reports*, **33** (1), 54–72.

128   Ongley, S.E., Bian, X., Neilan, B.A., Müller, R. (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Natural Product Reports*, **30** (8), 1121–1138.

129   Stevens, D.C., Hari, T.P.A., Boddy, C.N. (2013) The role of transcription in heterologous expression of polyketides in bacterial hosts. *Natural Product Reports*, **30** (11), 1391–1411.

130   Zhang, M.M., Wang, Y., Ang, E.L., Zhao, H. (2016) Engineering microbial hosts for production of bacterial natural products. *Natural Product Reports*, 33 (8), 963-987.

131   Engler, C., Kandzia, R., Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, **3** (11), e3647.

132   Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A.3., Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, **6** (5), 343–345.

133   Zhang, Y., Buchholz, F., Muyrers, J.P., Stewart, A.F. (1998) A new logic for DNA engineering using recombination in *Escherichia coli. Nature Genetics*, **20** (2), 123–128.

134   Zhang, Y., Muyrers, J.P., Testa, G., Stewart, A.F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nature Biotechnology*, **18** (12), 1314–1317.

135   Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F., Zhang, Y. (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature Biotechnology*, **30** (5), 440–446.

136   Wang, H., Li, Z., Jia, R., Hou, Y., Yin, J., Bian, X., Li, A., Müller, R., Stewart, A.F., Fu, J., Zhang, Y. (2016) RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nature Protocols*, **11** (7), 1175–1190.

137   Noskov, V., Kouprina, N., Leem, S.-H., Koriabine, M., Barrett, J.C., Larionov, V. (2002) A genetic system for direct selection of gene-positive clones during recombinational cloning in yeast. *Nucleic Acids Research*, **30** (2), E8.

138   Kouprina, N., Noskov, V.N., Koriabine, M., Leem, S.-H., Larionov, V. (2004) Exploring transformation-associated recombination cloning for selective isolation of genomic regions. *Methods in Molecular Biology (Clifton, N.J.)*, **255**, 69–89.

139   Wang, H., Bian, X., Xia, L., Ding, X., Müller, R., Zhang, Y., Fu, J., Stewart, A.F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection. *Nucleic Acids Research*, **42** (5), e37.

140   Gaj, T., Gersbach, C.A., Barbas, C.F.3. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, **31** (7), 397–405.

141   Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews. Molecular Cell Biology*, **14** (1), 49–55.

142   Le Cong, Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, **339** (6121), 819–823.

143   Julien, B. and Shah, S. (2002) Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*. *Antimicrobial Agents and Chemotherapy*, **46** (9), 2772–2778.

144   Perlova, O., Fu, J., Kuhlmann, S., Krug, D., Stewart, A.F., Zhang, Y., Müller, R. (2006) Reconstitution of the myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in

Myxococcus xanthus. Applied and Environmental Microbiology, **72** (12), 7485–7494.

145 Fu, J., Wenzel, S.C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A.F., Müller, R., Zhang, Y. (2008) Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Research*, **36** (17), e113.

146 Chai, Y., Shan, S., Weissman, K.J., Hu, S., Zhang, Y., Müller, R. (2012) Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. *Chemistry & Biology*, **19** (3), 361–371.

147 Tu, Q., Herrmann, J., Hu, S., Raju, R., Bian, X., Zhang, Y., Müller, R. (2016) Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. *Scientific Reports*, **6**, 21066.

148 Tang, L., Chung, L., Carney, J.R., Starks, C.M., Licari, P., Katz, L. (2005) Generation of new epothilones by genetic engineering of a polyketide synthase in *Myxococcus xanthus*. *The Journal of Antibiotics*, **58** (3), 178–184.

149 Kao, C.M., Katz, L., Khosla, C. (1994) Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science (New York, N.Y.)*, **265** (5171), 509–512.

150 Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., Julien, B. (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science (New York, N.Y.)*, **287** (5453), 640–642.

151 Pfeifer, B.A. and Khosla, C. (2001) Biosynthesis of polyketides in heterologous hosts. *Microbiology and Bolecular Biology Reviews: MMBR*, **65** (1), 106–118.

152 Schumann, J. and Hertweck, C. (2006) Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *Journal of Biotechnology*, **124** (4), 690–703.

153 Fujii, I. (2009) Heterologous expression systems for polyketide synthases. *Natural Product Reports*, **26** (2), 155–169.

154 Loeschcke, A. and Thies, S. (2015) *Pseudomonas putida*-a versatile host for the production of natural products. *Applied Microbiology and Biotechnology*, **99** (15), 6197–6214.

155 Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science (New York, N.Y.)*, **291** (5509), 1790–1792.

156 Wenzel, S.C. and Müller, R. (2005) Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Current Opinion in Biotechnology*, **16** (6), 594–606.

157 Mutka, S.C., Carney, J.R., Liu, Y., Kennedy, J. (2006) Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry*, **45** (4), 1321–1330.

158 Watanabe, K., Hotta, K., Praseuth, A.P., Koketsu, K., Migita, A., Boddy, C.N., Wang, C.C., Oguri, H., Oikawa, H. (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nature Chemical Biology*, **2** (8), 423–428.

159 Gao, X., Wang, P., Tang, Y. (2010) Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Applied Microbiology and Biotechnology*, **88** (6), 1233–1242.

160 Yuzawa, S., Kim, W., Katz, L., Keasling, J.D. (2012) Heterologous production of polyketides by modular type I polyketide synthases in *Escherichia coli*. *Current Opinion in Biotechnology*, **23** (5), 727–735.

161 Antosch, J., Schaefers, F., Gulder, T.A.M. (2014) Heterologous Reconstitution of Ikarugamycin Biosynthesis in *E. coli*. *Angewandte Chemie International Edition*, **53** (11), 3011–3014.

162 Stewart, E.J. (2012) Growing unculturable bacteria. *Journal of Bacteriology*, **194** (16), 4151–4160.

163 Gray A.N., Koo B.M., Shiver A.L., Peters J.M., Osadnik H., Gross C.A. (2015) High-throughput bacterial functional genomics in the sequencing era. *Current Opinion in Microbiology*, **27**:86-95.

164 Garmendia L., Hernandez A., Sanchez M.B., Martinez J.L. (2012) Metagenomics and antibiotics. *Clinical Microbiology and Infection*, **18** (Suppl. 4):27-31.

165 Zhou J., He Z., Yang Y., Deng Y, Tringe S.G., Alvarez-Cohen L. High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *MBio*, **6**(1): e02288-14.

166 Gawad C., Koh W., Quake S.R. Single-cell genome sequencing: current state of the science. *Nature Review Genetics*, **17**(3):175-188.

167 Weissman, K.J. and Müller, R. (2009) A brief tour of myxobacterial secondary metabolism. *Bioorganic & Medicinal Chemistry*, **17** (6), 2121–2136.

168   Reichenbach, H. and Höfle, G. (1999) Myxobacteria as producers of secondary metabolites, in *Drug Discovery from Nature* (eds S. Grabley and R. Thiericke), Springer, Berlin, pp. 149–179.

169   Koch, A.L. and White, D. (1998) The social lifestyle of myxobacteria. *Bioessays*, **20** (12), 1030–1038.

170   Velicer, G.J. and Vos, M. (2009) Sociobiology of the myxobacteria. *Annual Review of Microbiology*, **63**, 599–623.

171   Cao, P., Dey, A., Vassallo, C.N., Wall, D. (2015) How Myxobacteria Cooperate. *Journal of Molecular* Biology, **427**(23):3709-3721.

172   Berleman, J. and Keane, R. (2016) The predatory life cycle of *Myxococcus xanthus*. *Microbiology*, **162** (1), 1–11.

173   Dworkin, M. (2008) Lingering Puzzles about Myxobacteria. *Microbe*, **2** (1), 18–24.

174   Wenzel, S.C. and Müller, R. (2007) Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Natural Product Reports*, **24** (6), 1211–1224.

175   Weissman, K.J. and Müller, R. (2010) Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Natural Product Reports*, **27** (9), 1276–1295.

176   Schäberle, T.F., Lohr, F., Schmitz, A., König, G.M. (2014) Antibiotics from myxobacteria. *Natural Product Reports*, **31**(7):953-72.

177 Herrmann J., Fayad A.A., Müller R. (2016) Natural products from myxobacteria: novel metabolites and bioactivities. *Natural Product Reports,* DOI:10.1039/c6np00106h

178   Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J.A., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H.S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S.A., Vaudin, M., Wiegand, R., Kaplan, H.B. (2006) Evolution of sensory complexity recorded in a myxobacterial genome. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (41), 15200–15205.

179   Wenzel, S.C. and Müller, R. (2009) Myxobacteria—'microbial factories' for the production of bioactive secondary metabolites. *Molecular BioSystems*, **5** (6), 567–574.

180   Wenzel, S.C. and Müller, R. (2009) The biosynthetic potential of myxobacteria and their impact on drug discovery. *Current Opinion in Drug Discovery & Development*, **12** (2), 220–230.

181   Wenzel, S.C. and Müller, R. (2009) The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Natural Product Reports*, **26** (11), 1385–1407.

182   Wenzel, S.C. and Müller, R. (2010) Myxobacteria - unique microbial secondary metabolite factories, in *Comprehensive Natural Products Chemistry II, Vol 2: Structural Diversity II - Secondary Metabolite Sources, Evolution and Selected Molecular Structures* (ed B. Moore), Elsevier, Oxford, pp. 189–222.

183   Buntin, K., Irschik, H., Weissman, K.J., Luxenburger, E., Blocker, H., Müller, R. (2010) Biosynthesis of thuggacins in myxobacteria: comparative cluster analysis reveals basis for natural product structural diversity. *Chemistry & Biology*, **17** (4), 342–356.

184   Kopp, M., Irschik, H., Gemperlein, K., Buntin, K., Meiser, P., Weissman, K.J., Bode, H.B., Müller, R. (2011) Insights into the complex biosynthesis of the leupyrrins in *Sorangium cellulosum* So ce690. *Molecular BioSystems*, **7** (5), 1549–1563.

185   Pistorius, D. and Müller, R. (2012) Discovery of the rhizopodin biosynthetic gene cluster in *Stigmatella aurantiaca* Sg a15 by genome mining. *Chembiochem: a European Journal of Chemical Biology*, **13** (3), 416–426.

186   Hoffmann, T., Müller, S., Nadmid, S., Garcia, R., Müller, R. (2013) Microsclerodermins from terrestrial myxobacteria: an intriguing biosynthesis likely connected to a sponge symbiont. *Journal of the American Chemical Society*, **135** (45), 16904–16911.

187   Sucipto, H., Wenzel, S.C., Müller, R. (2013) Exploring chemical diversity of alpha-pyrone antibiotics: molecular basis of myxopyronin biosynthesis. *Chembiochem: a European Journal of Chemical Biology*, **14** (13), 1581–1589.

188   Bouhired, S.M., Crusemann, M., Almeida, C., Weber, T., Piel, J., Schaberle, T.F., Konig, G.M. (2014) Biosynthesis of phenylnannolone A, a multidrug resistance reversal agent from the halotolerant myxobacterium *Nannocystis pusilla* B150. *Chembiochem: a European Journal of Chemical Biology*, **15** (5), 757–765.

189   Osswald, C., Zaburannyi, N., Burgard, C., Hoffmann, T., Wenzel, S.C., Müller, R. (2014) A highly unusual polyketide synthase directs dawenol polyene biosynthesis in *Stigmatella aurantiaca*. *Journal of Biotechnology*, **191**, 54–63.

190   Keller, L., Plaza, A., Dubiella, C., Groll, M., Kaiser, M., Müller, R. (2015) Macyranones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *Journal of the American Chemical Society*, **137** (25), 8121–8130.

191   Wenzel, S.C., Hoffmann, H., Zhang, J., Debussche, L., Haag-Richter, S., Kurz, M., Nardi, F., Lukat, P., Kochems, I., Tietgen, H., Schummer, D., Nicolas, J.-P., Calvet, L., Czepczor, V., Vrignaud, P., Muhlenweg, A., Pelzer, S., Müller, R., Bronstrup, M. (2015) Production of the Bengamide Class of Marine Natural Products in Myxobacteria: Biosynthesis and Structure-Activity Relationships. *Angewandte Chemie (International ed. in English)*, **54** (51), 15560–15564.

192   Sun, Y., Feng, Z., Tomura, T., Suzuki, A., Miyano, S., Tsuge, T., Mori, H., Suh, J.-W., Iizuka, T., Fudou, R., Ojika, M. (2016) Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Scientific Reports*, **6**, 22091.

193   Müller, S., Rachid, S., Hoffmann, T., Surup, F., Volz, C., Zaburannyi, N., Müller, R. (2014) Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains. *Chemistry & Biology*, **21** (7), 855–865.

194   Erol, O., Schaberle, T.F., Schmitz, A., Rachid, S., Gurgui, C., El Omari, M., Lohr, F., Kehraus, S., Piel, J., Müller, R., Konig, G.M. (2010) Biosynthesis of the myxobacterial antibiotic corallopyronin A. *ChemBioChem: a European journal of chemical biology*, **11** (9), 1253–1265.

195   Irschik, H., Kopp, M., Weissman, K.J., Buntin, K., Piel, J., Müller, R. (2010) Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. *Chembiochem: a European Journal of Chemical Biology*, **11** (13), 1840–1849.

196   Pistorius, D., Li, Y., Sandmann, A., Müller, R. (2011) Completing the puzzle of aurachin biosynthesis in *Stigmatella aurantiaca* Sg a15. *Molecular BioSystems*, **7** (12), 3308–3315.

197   Jungmann, K., Jansen, R., Gerth, K., Huch, V., Krug, D., Fenical, W., Müller, R. (2015) Two of a Kind—The Biosynthetic Pathways of Chlorotonil and Anthracimycin. *ACS Chemical Biology*, **10** (11), 2480–2490.

198   Young, J., Stevens, D.C., Carmichael, R., Tan, J., Rachid, S., Boddy, C.N., Müller, R., Taylor, R.E. (2013) Elucidation of gephyronic acid biosynthetic pathway revealed unexpected SAM-dependent methylations. *Journal of Natural Products*, **76** (12), 2269–2276.

199   Schummer, D., Höfle, G., Forche, E., Reichenbach, H., Wray, V., Domke, T. (1996) Antibiotics from Gliding Bacteria, LXXVI. Vioprolides: New Antifungal and Cytotoxic Peptolides from *Cystobacter violaceus*. *Liebigs Annalen-Recueil*, **1996** (6), 971–978.

200   Chopin, N., Couty, F., Evano, G. (2010) Synthesis of the Azetidinyl-Thiazoline Fragment of Vioprolides A and C. *Letters in Organic Chemistry*, **7** (5), 353–359.

201   Liu, H. and Thomas, E.J. (2013) Synthesis of the (*E*)-dehydrobutyrine–thiazoline–proline–leucine fragment of vioprolides B and D. *Tetrahedron Letters*, **54** (24), 3150–3153.

2. MANUSCRIPT

Biosynthesis and heterologous expression of vioprolide gene cluster revealing a C domain catalyzed glycerate esterification and a post-assembly maturation

Fu Yan‡, David Auerbach‡, Lena Keller, Qiang Tu, Youming Zhang*, Rolf Müller*

Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Center for Infection Research and Pharmaceutical Biotechnology, Saarland University Campus Building E8.1, 66123 Saarbrücken (Germany), E-mail: Rolf.Mueller@helmholtz-hzi.de.

Shandong University – Helmholtz Joint Institute of Biotechnology, State Key Laboratory of Microbial Technology, School of Life Science, Shandong University, Jinan 250100, People's Republic of China, E-mail: zhangyouming@sdu.edu.cn.

KEYWORDS: hydrolytic maturation, NRPS, vioprolides, glycerate esterification

Author's contributions:

Fu Yan designed the research, prepared the manuscript; performed sequence analysis, genetic engineering, compounds purification, designed and performed *in vitro* biochemical experiments.

David Auerbach prepared the manuscript, designed and performed *in vitro* biochemical experiments, purified compounds and elucidated chemical structures.

Yi Chai cloned the *vio* gene cluster and performed genetic engineering.

Lena Keller elucidated structures.

Qiang Tu performed genetic engineering.

Youming Zhang designed and supervised the research.

Rolf Müller designed and supervised the research.

2. Manuscript

Abstract

Vioprolides, produced by *Cystobacter violaceus* Cb vi35, are myxobacterial peptolides notable for their prominent cytotoxicity and anti-fungal activity. While their structures were reported back in 1996, the biosynthesis, in particular that of the uncommon structural element 4-methylazetidine-carboxylic acid (Maz) remained elusive. We successfully established heterologous expression systems in *Myxococcous xanthus* DK1622 (ΔmchA), *Burkholderia sp.* DSM7029 and *Pseudomonas putida* KT2440 for the vioprolide biosynthetic gene cluster which enabled us to study the biosynthetic machinery. Gene inactivation studies in these hosts allowed us to identify the genes involved in vioprolide biosynthesis and the gene responsible for the formation of pipecolic acid which is indispensable for the biosynthesis of vioprolides A and B. A discrepancy between the biosynthetic gene cluster and the known vioprolide structures was revealed as a hydrolytic maturation process. Vioprolides were found to be synthesized as acylated precursors by nonribosomal peptide synthetase (NRPS) machinery containing an unusual starter domain. After assembly, these acylated precursors were cleaved and only the matured vioprolides were exported into the extracellular medium. An unusual C domain was identified to catalyze the condensation between fatty acids and glycerate, which was characterized by *in vitro* biochemical analysis.

2. Manuscript

2.1 Introduction

Vioprolids are striking examples for the structural diversity that can be found in myxobacterial secondary metabolites (Fig. 2.1). The peptolide produced by the myxobacterium *Cystobacter violaceus* Cb vi35 are one of the very few examples of bacterial metabolites containing an azetidine ring and are interesting compounds for further research due to their remarkable cytotoxic, antifungal and immunomodulatory activities (US patents: US20100028298 and US20110173707) [1]. In view of the exerted effects, it seems likely that several independent targets are affected by vioprolides and the combination of anti-fungal and immunomodulatory activity is noteworthy. Fungal infections receive less attention than the more threatening bacterial infections in addition to the availability of limited data regarding the epidemiology of fungi related infections [2]. However, this data suggests that this issue is underestimated and that the growing number of immunocompromised patients is highly threatened by invasive mycoses [3]. We think that vioprolide D is a promising compound following a one-compound-multiple-targets strategy for invasive fungal infections, since it acts as an anti-fungal agent in addition to inhibiting interferon response. This combination is noteworthy, since it has been shown that type I interferon-signaling mediates the lethal hyper-inflammatory response during systemic mouse infections with *C. albicans* [4]. To further elaborate on the mode and the mechanism of action, a steady supply of compound is necessary.

In this study we established several heterologous hosts successfully producing vioprolides with one host producing at significantly higher titers than the natural producer. The genomic data available for Cb vi35 allowed us to identify the biosynthetic gene cluster. The biosynthetic model for vioprolide formation was in discrepancy with the reported structures as the presence of a fatty acid ligase and a bifunctional glyceryl transferase/phosphatase belonging to the HAD superfamiliy was found to initiate the biosynthesis. This contradiction led to the

identification of acylated vioprolides that were exclusively found in the cell pellet. Next to the cystomanamides, this is to our knowledge the only other example of an FkbH-like domain in an NRPS module. The acyl group that is introduced by this unusual initiation module likely represents another example of a hydrolytic maturation process.
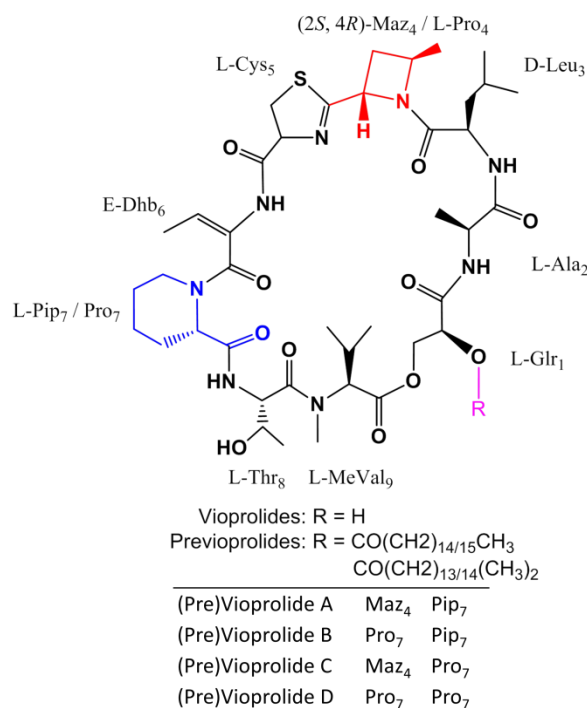
Fig. 2.1 Chemical structures of vioprolides

2.2 Results & Discussion

Identification of acylated vioprolides

In the course of our efforts to shed light on the biosynthesis of vioprolides, we identified a fatty acid ligase in the biosynthetic cluster that is in discrepancy with the known structures of vioprolides A-D. To identify these putative acyl derivatives, we screened data obtained in feeding studies of *C. violaceus* Cb vi35 for the presence of compounds that also exhibited the corresponding mass shifts. Interestingly, no such compounds could be found in the culture supernatants, but in the cell pellets (supplementary). The incorporation of

L-methionine-(*methyl*-^{13}C), L-methionine-^{13}C$_4$,^{15}N, L-methionine-(*methyl*-d$_3$) and L-threonine-^{13}C$_4$,^{15}N was observed for a range of four compounds with retention times shifted to high acetonitrile content (supplementary). These compounds were subsequently assigned based on exact mass and fragmentation pattern as vioprolide A/B and vioprolide C/D acylated derivatives that were not known before. The successful purification and structure elucidation of these derivatives was only possible after establishing a heterologous expression system that enabled us to isolate a sufficient amount of compound. The presence of the acyl derivatives was found in agreement with the biosynthetic gene cluster found in *C. violaceus* Cb vi35, details regarding the biosynthesis will be given in the next paragraph. No fatty acids other than saturated C$_{16}$ and C$_{17}$ were found in the natural producer as well as the heterologous hosts which is in agreement with reported lipid profiles of myxobacteria [5–7]. Four precursors could be purified from 10 L cultures of *M. xanthus*::Ptet-vio with yields between 1.2 mg and 11.9 mg. The most prominent precursor was hydrolyzed by sodium hydroxide and the fatty acyl component was identified as *iso*-C$_{17:0}$ by GC-MS (supplementary). Structure elucidation by 1D and 2D NMR experiments revealed the structure as vioprolide B with an *iso*-C17 fatty acid chain on the free hydroxyl group of glyceric acid (supplementary). The bioactivity assay showed that previoprolide B retained the high cytotoxicity on HCT-116 colon cancer cells (IC$_{50}$ = 0.12 μg/mL), while no obvious activity against *Candida albicans* and *Mucor hiemalis* (IC$_{50}$ > 64 μg/mL) was found, which was consistent with the bioactivity of vioprolide B. The fact that no obvious change in bioactivity could be found for the acylated vioprolide suggests the acylation does not serve as a pro-drug mechanism but rather serves a directed transport mechanism, since acylated variants were exclusive to the cell pellet.

Identification and sequence analysis of the *vio* gene cluster

The genome sequence of *C. violaceus* Cb vi35 was searched *in silico* for secondary metabolite gene clusters using the antiSMASH platform revealing a region of 56.2 kb (68.9% GC content) that, based on retro-biosynthetic considerations, putatively contains the vioprolide gene cluster [8, 9]. The core biosynthetic assembly line contains 10 modules harboring 33 domains (Fig. 2.2) and is flanked by a putative ornithine cyclodeamidase (*vioZ*) upstream of *vioA* and a putative biosynthetic protein (*orf3*) downstream of *vioD*. The biosynthesis is driven by 2 unusual modules and 8 NRPS modules encoded on the four genes *vioA-vioD*. After antiSMASH analysis, annotations were further refined by alignment with the non-redundant protein database. The resulting *in silico* based assignments are summarized in (supplementary). Modules 3, 4, 6, 9 and 10 are standard NRPS-modules for which *in silico* analysis of A domain specificities is in agreement with the incorporated amino acids [10]. The A_3 domain activates L-Pro and probably also 4-methylazetidine carboxylic acid (Maz). Its binding pocket (DVQCLSEVTK) is different from A_{Pro} domains (DVQLIAHVVK). Besides, several active residues of the Pip/Pro activating (A_6) domain (DIQYYAQVVK) are also different from reported Pip or Pro activating domain. The non-standard modules were subject to a detailed analysis. The identified cluster was heterologously expressed in three different hosts for production optimization and gene knock-out experiments. Knock-out experiments were carried out in *Myxococcus xanthus* DK1622 and *Burkholderia sp.* DSM7029 as a phylogenetically distant host. Ancillary genes *orf3 – orf9* were knocked out one by one as well as all at once, however no effect on the production of vioprolides in both heterologous hosts was observed and these genes were thus deemed not involved in the biosynthesis of vioprolides (supplementary).
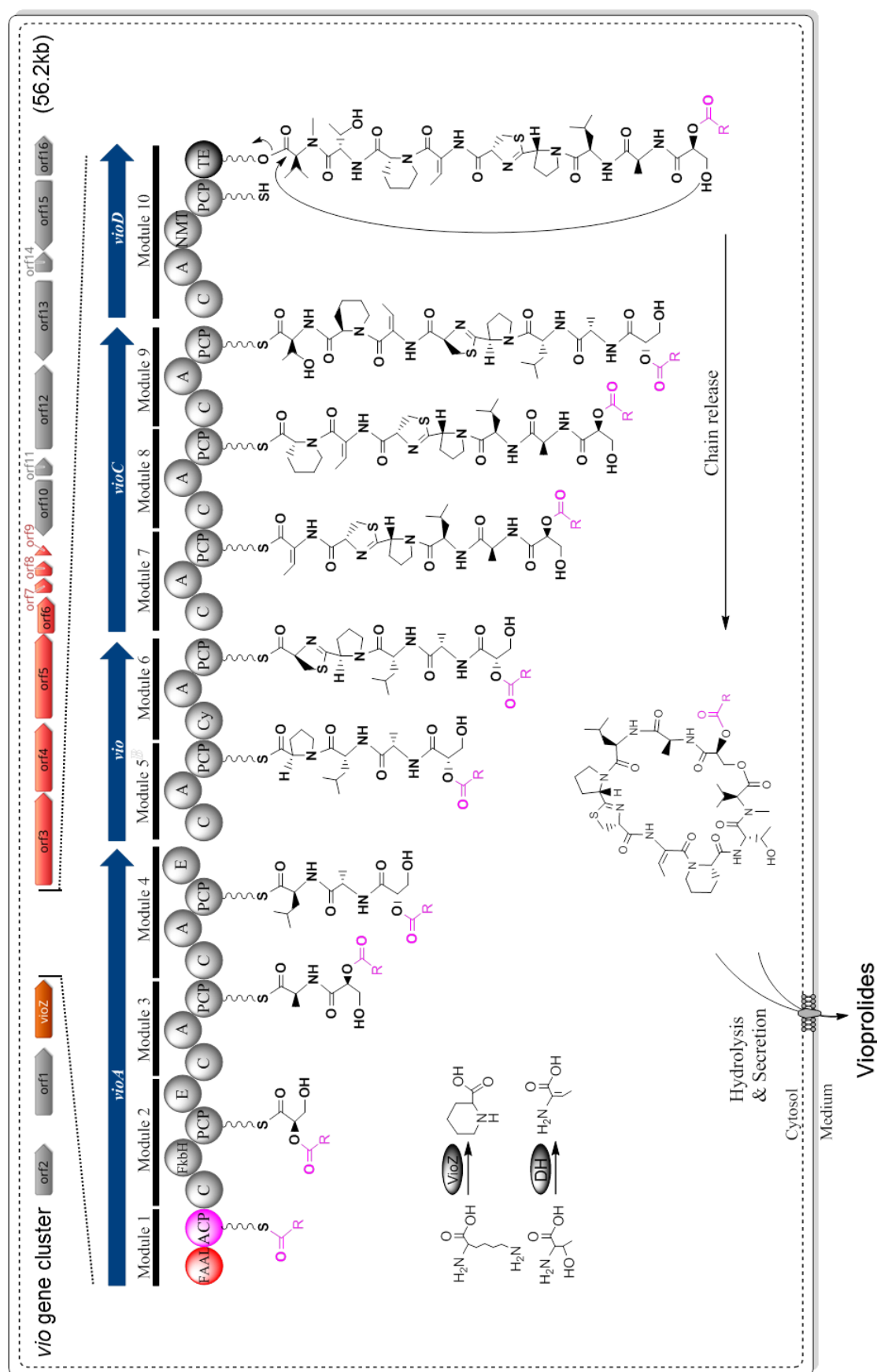
Fig. 2.2 Diagram of *vio* gene cluster and the proposed biosynthetic pathway of vioprolides.

In vitro **characterization of module 1**

The initiating module starts with a FAAL (fatty acid AMP ligase) domain which was characterized to activate and introduce long chain fatty acids in both polyketide and nonribosomal peptide biosynthesis by activation and loading of fatty acids to ACP domains [11–13]. Since the known vioprolides do not contain fatty acids, this led us to believe that the acylation is an intermediate means to facilitate compound export. Several examples of hydrolytic maturation in secondary metabolite biosynthesis were recently reported. In the biosynthesis of xenocoumarin [14], heronamide [15], peanilamicin [16], pyoverdine [17], colibactin [18], naphthyridinomycin [19], quinocarcin [20] and telomycin [21] a hydrolytic removal of *N*-amino-acyl chains was shown to serve as a maturation process that was also responsible for the observed discrepancy between the biosynthetic gene cluster and the structure of the corresponding compounds in some cases. Proteolytic maturation was also shown for didemnin [22], saframycin [23] and nocardicin A [24] biosynthesis but the corresponding proteases are unknown. A comprehensive overview of maturation processes in natural compounds biosynthesis can be found in recent reviews [25, 26].

In order to validate the C domain catalyzed glycerate esterification, the full-length FAAL-ACP-C-FkbH-PCP (FkbH-PCP, 205 kDa), FAAL-ACP, and ACP were expressed and purified from *E. coli* (supplementary). FAAL-ACP was initially incubated with MtaA (PPTase) and fatty acids. The reaction without MtaA, ATP or fatty acid served as a negative control. To our surprise, we only found the loading of palmitic acid in the presence of other fatty acids and even when no fatty acid was added to the assay (supplementary). Once *apo*-FAAL-ACP was transformed to *holo*-FAAL-ACP by incubation with MtaA and ATP, $C_{16:0}$ was transferred completely. Without the presence of ATP, the loading of palmitic acid was incomplete (supplementary). These results suggested the presence of an internal activated fatty acid bound to the FAAL domain and that

the fatty acyl loading is ATP-dependent. As previously reported, an acyl adenylate binding pocket exists in FAAL [23]. We hypothesized that FAAL activates palmitic acid, the most prominent fatty acid [27] in *E. coli*, and held it in its hydrophobic pocket. In order to get rid of this intrinsic fatty acid, we first attempted to purify soluble FAAL-ACP from inclusion bodies by denaturation and refolding but failed. In a second approach we attempted to chemically offload the fatty acid with cysteamine as reported by Belecki *et al.* to enable subsequent loading assays [28]. The palmitoyl chain could be successfully offloaded but at the same time excessive disulfide bridges formed between FAAL-ACP and cysteamine which could not be fully converted even after 3 hours of incubation at 37 °C with DTT. The third and successful strategy to obtain FAAL-ACP without innate fatty was an *in trans* catalytic approach. The binding pocket of the FAAL domain was initially emptied by transferring the internal palmitoyl chain to the downstream ACP of FAAL-ACP. The pocket vacant FAAL-ACP was then utilized for *in trans* transfer of $C_{16:0}$, $C_{17:0}$, *iso*-$C_{16:0}$ and *iso*-$C_{17:0}$ to the discrete *holo*-ACP. The fatty acids were restricted to these four species in consideration of the identified compounds. Characterization of the protein loadings by LC-MS revealed *iso*-$C_{17:0}$ as the favored substrate (Fig. 2.3). This result is consistent with the structure of previoprolide B which was identified as the most prominent precursor.
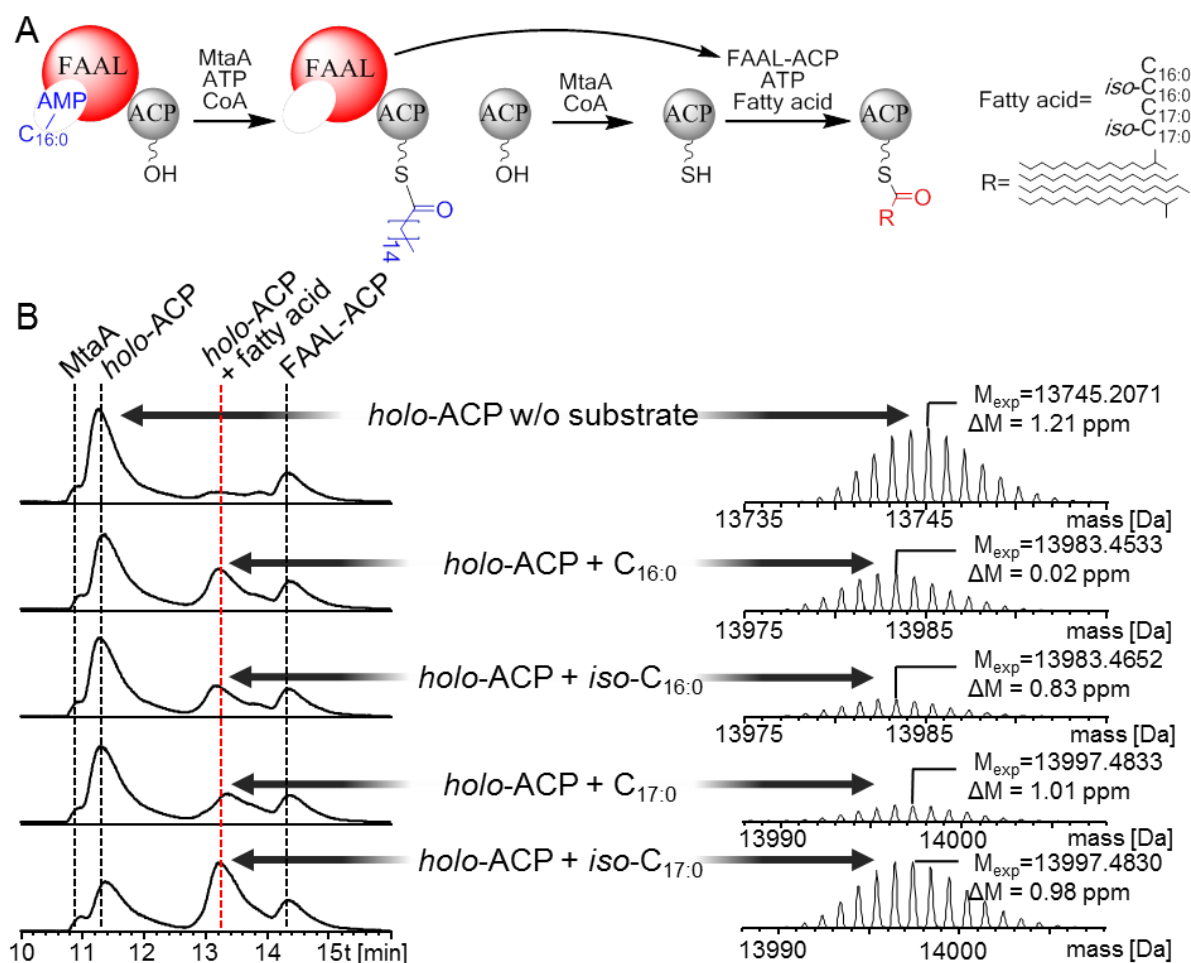
Fig. 2.3 Biochemical characterization of the module 1. (A) scheme of enzymatic reaction. The binding pocket of FAAL-ACP was emptied by incubation with MtaA, ATP and CoA. The pocket free FAAL-ACP was used to load other fatty acids to the discrete ACP; (B) LC-MS analysis of fatty acids loaded on *holo*-ACP.

In vitro characterization of module 1&2

The second domain of module 2 was annotated as FkbH-like domain belonging to the HAD-superfamily which is expected to load and dephosphorylate D-1, 3-bisphosphoglycerate to afford a glyceryl moiety on the PCP$_1$ domain. Such bifunctional glyceryl transferases/phosphatases were reported in other NRPS/PKS assembled natural products such as FK520, oxazolomycin, tautomycin, tetronomycin, RK-682, zwittermicin A, FR901464,

phormidolide, cystomanamide and pellasoren [29–38]. Most of the reported FkbH-like proteins serve to provide a variety of extender or starter units to PKS biosynthetic machinery with cystomanamide as the only example of an FkbH-like protein in a NRPS setting. To the best of our knowledge, this is the first example of an FkbH domain initiated NRPS assembly line, further underlining the generality of glyceric acid as a building block. In a detailed study of the OmzB protein by Dorrestein *et al.* it was shown that the first step involved in the process is the loading of D-1, 3-bisphosphoglycerate [31]. Vioprolides, however contain an L-configured glycerate moiety which fits well with the E_1 domain that putatively isomerizes D-glycerate to L-glycerate. An alignment of the C domains with other C domains (supplementary) led to a prediction that is contradicting the known structures. A possible reason for this apparent discrepancy is that the underlying data used for alignment is lacking the critical number of C domains with the same substrate specificity. Validation of the C domain catalyzed glycerate esterification was initially attempted on the intact protein level. *In vitro* reconstitution assay were done with heterologously expressed FAAL-ACP-C-FkbH-PCP (**1**), FAAL-ACP (**2**), C-FkbH-PCP (**3**), FkbH-PCP (**4**) and PCP (**5**). *In cis* assays using **1** could not be successfully established due to the instability of **1** under various buffer and reaction conditions. Combinations of **1** and **3**, **1** and **4**, **1** and **5,** and **2** and **3** were used for *in trans* catalysis, however no fatty acyl glycerate transfer to the *holo*-PCP could be detected. To substantiate the putative condensation of the fatty acid with glyceric acid, we decided to characterize the enzyme bound substrates by making use of the offloading strategy reported by Belecki *et al.* Briefly, protein samples of **1** were incubated with D-3-phosphoglyceric acid , phosphoglycerate kinase (PGK), MtaA, ATP, CoA and fatty acid. After this step, samples were further incubated with cysteamine at 4 °C overnight to promote the chemical release of the bound substrates which were then extracted and measured by LC-MS/MS. Since it was

known that $C_{16:0}$ was inherent to the FAAL domain, the incorporation of fatty acids other than palmitic acid relied on a certain turnover once cysteamine was added and thus also relied on the presence of PGK. FAAL bound fatty acids were found as single compounds while two species with different retention times were found for each fatty acyl glyceryl derivative with identical MS/MS fragmentation pattern and accurate mass (Fig. 2.4). These two species are likely the regioisomers formed via 2, 3-acyl migration – a process that is reported to facilely take place under various experimental conditions [39–42]. We also found the $C_{16:0}$ acylated glyceric acid derivative in the absence of all enzymes but **1** indicating an almost complete attachment of glycerate in *E. coli*. As expected, the presence of glyceric acid derivatives of $C_{17:0}$ and *iso*-$C_{17:0}$ could only be detected in the presence of PGK and phosphoglycerate (Fig. 2.4).
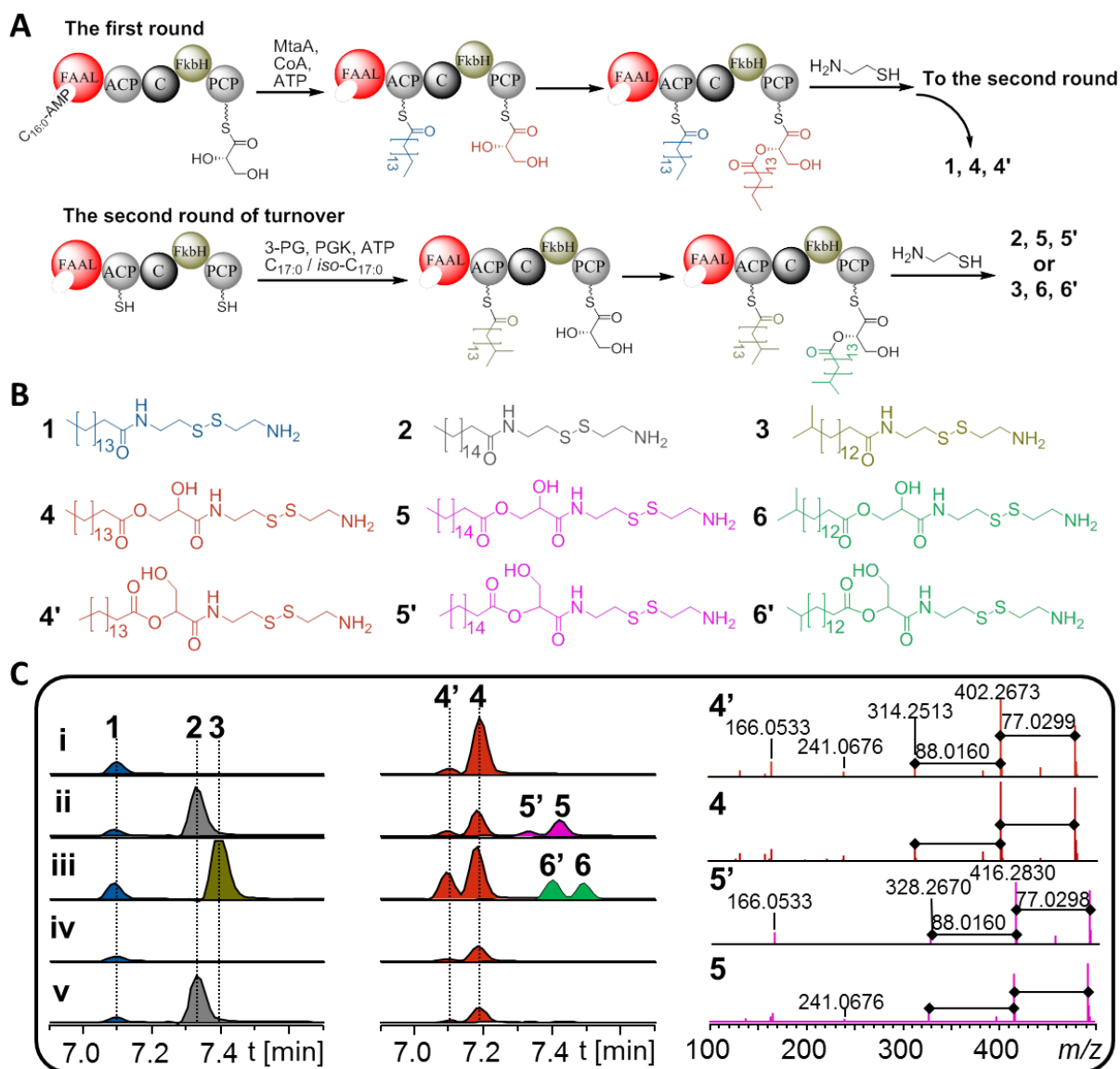
Fig. 2.4 Biochemical characterization of module 1&2. (A) diagram of offloading intermediates from FAAL-PCP; in the first round of reaction, the innate $C_{16:0}$ was transferred to the ACP and was subsequently condensed with glycerate; after offloading $C_{16:0}$-glycerate and the inherent $C_{16:0}$, the net FAAL-PCP went to the second round of reaction with $C_{17:0}$ or *iso*-$C_{17:0}$ and reacted in a similar way; (B) structures of the offloaded protein-bound intermediates in various reaction systems; (C) UPLC-HRMS analysis of the offloaded intermediates; reaction systems: i) FAAL-PCP control without other enzymes and substrates; ii) FAAL-PCP + $C_{17:0}$ + MtaA + ATP + PGK; iii) FAAL-PCP + *iso*-$C_{17:0}$ + MtaA + ATP + PGK; iv) FAAL-PCP + MtaA + ATP + PGK (without fatty acid); v) FAAL-PCP + *iso*-$C_{17:0}$ + MtaA + ATP (without PGK); the MS/MS patterns of 4 and 5 as well as their regioisomers (4' and 5' respectively) are shown; mass accuracy is below 100 ppb.

Overexpression of the *vio* gene cluster in heterologous hosts

The myxobacterium *Myxococcus xanthus* served as heterologous host to produce several compounds as reported previously [43–49]. The heterologous host *M. xanthus* DK1622 (*ΔmchA*) used in this study is derived from *M. xanthus* DK1622 in which the native myxochromide A gene cluster was deleted. After integration of p15A-Ptet-vio into the chromosome of *M. xanthus* DK1622 (*ΔmchA*), several transformants *M. xanthus*::Ptet-vio were cultivated in CTT medium to test the production of vioprolides. The UHPLC-ESI-HRMS analysis revealed the production of four major compounds. The first compound showed the same retention time, molecular mass and MS/MS fragmentation patterns as vioprolide D, and the third compound was consistent with vioprolide A/B (Fig. 2.5). However, the retention time of the second and the fourth compounds did not match the reported vioprolides, although their molecular mass and MS/MS fragmentation patterns are identical with vioprolide C/D and A/B, respectively. These compounds were subsequently purified and validated to be vioprolides by NMR (supplementary). The first and the third compounds turned out to be vioprolide D and B, while the second and the fourth compounds are the regionisomers of vioprolide D and B. Quantitative analysis showed the production yield of vioprolide B (39.6 mg/L) and D (41.6 mg/L) was 6-fold and 9-fold with half producing time as in the native producer (supplementary). In addition, considering *Ptn5* promoter significantly improved the production of myxochromide S in *M. xanthus* [45], we replaced the *Ptet* promoter in p15A-Ptet-vio with *Ptn5* promoter to study whether the production titer of vioprolides could be elevated. The resulting expression vector p15A-Ptn5-vio was transformed into *M. xanthus* DK1622 (*ΔmchA*). The UHPLC-ESI-HRMS analysis showed that all vioprolides were produced in the mutant *M. xanthus*::Ptn5-vio, however, the production yield of vioprolide B (10.0 mg/L) and D (6.7 mg/L) was lower than *M. xanthus*::Ptet-vio.

Burhkolderia is a Gram-negative genus revealing great potential in natural products biosynthesis [50–52]. Here we also studied the feasibility to express the *vio* gene cluster in *Burkholderia*. Both p15A-Ptet-vio and p15A-Ptn5-vio were integrated into the genome of *Burkholderia sp.* K481-B101 (ATCC 53080; DSM 7029) [53, 54]. The transformants *B. sp.* DSM7029::Ptet-vio and *B. sp.* DSM7029::Ptn5-vio were fermented in CYCG medium. UHPLC-ESI-HRMS analysis revealed the production of vioprolide B and D (Fig. 2.5), although only vioprolide B was the prominent product. Quantitative analysis showed that *B. sp.* DSM7029::Ptet-vio and *B. sp.* DSM7029::Ptn5-vio produced 16.9 mg/L and 20.4 mg/L vioprolide B, respectively. Our results suggest the great potential of *Burkholderia* to produce secondary metabolites as heterologous host. The vioprolide expression vectors were also introduced into *Pseudomonas putida* KT2440 which is widely used as heterologous host to produce natural products [53], but only low amounts of vioprolide B and D were produced (supplementary). In addition, p15A-Ptet-vio was transformed into *E. coli* GB2005 (MtaA) [56] and *E. coli* Nissle 1917 [54], but no vioprolide production was observed in various conditions (data not shown).
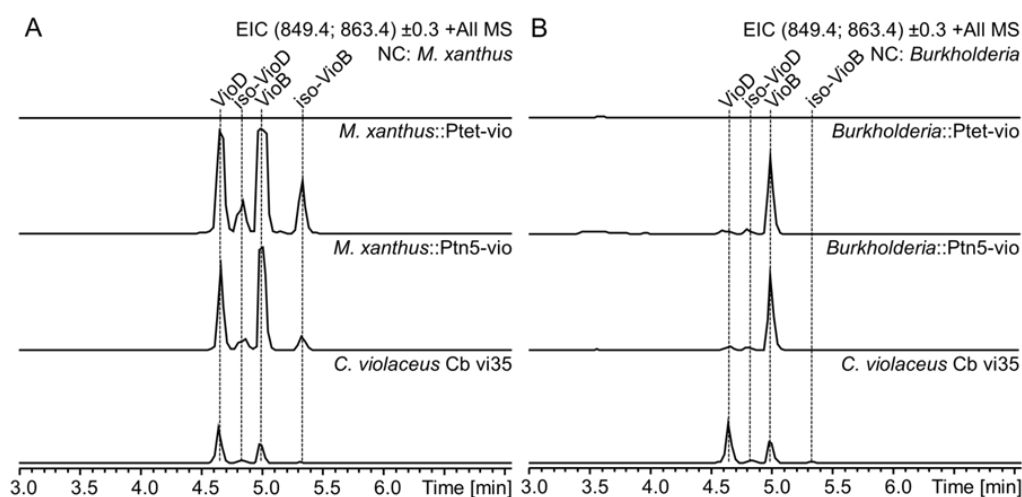

Fig. 2.5 UHPLC-ESI-HRMS analysis of vioprolides from *M. xanthus* and *Burkholderia* mutants. Extracted ion chromatogram (EIC) *m/z* of vioprolides (vioprolide D and isomer, $[M+H]^+ = 849.4$; vioprolide B and isomer, $[M+H]^+ = 863.4$) were analyzed for *M. xanthus* DK1622 (*ΔmchA*) mutants (A) and *B. sp.*

DSM7029 mutants (B). *C. violaceus* Cb vi35 was set as reference, *M. xanthus* and *B. sp* DSM7029 without *vio* gene cluster were set as negative control (NC). Intensity is set to the same range.

Mechanism of Maz formation

Only several natural products containing azetidine rings are known, and even less information can be found regarding the biosynthetic origin of the four-membered heterocycle. Among the plant derived compounds mugineic acid [57], calydaphninone [58], gelsemoxonine [59] and azetidine-2-carboxylic acid, biosynthesis studies showed that azetidine is formed from 2, 4-diaminobutyric acid which itself originates from *S*−adenosylmethionine (SAM) or homoserine [60]. The biosynthesis of the sponge derived compound penaresidin [61] and the fungal metabolites okaramine B [62] and taichunamide [63] is completely unknown. The azetidine ring of polyoxins produced by *Streptomyces cacaoi* is formed by heterocyclization of isoleucine [64–66]. We performed a BLAST search of the three genes *polC*, *polE* and *polF* that were suspected to be responsible for azetidine formation in polyoxins, but no matches were found in the genomes of *C. violaceus* Cb vi35 and *M. xanthus* DK1622. The four-membered azetidine ring in penicillin and clavulanic acid was formed via intramolecular cyclization catalyzed by dedicated synthase [67], whereas the β-lactam in nocardicin A was formed by an unusual C domain catalyzed intermolecular cyclization [68]. Unexpectedly, *C. violaceus* Cb vi35 didn't produce vioprolide A and C under the conditions applied in our experiment, so it is not possible to identify the precursor of Maz by isotope feeding. It seems that certain conditions, e.g. chemical signals from environment, are essential to activate the expression of Maz forming enzymes. We are trying to cultivate *C. violaceus* Cb vi35 in various conditions to get the production of vioprolide A and C. As vioprolide A and C were not produced in heterologous hosts, the Maz

forming genes may not exist in the cloned *vio* gene cluster. It is also possible that certain cultivating condition is needed to activate the expression of Maz forming enzymes in heterologous systems. Therefore, the mechanism of Maz formation is still elusive at present.

Generation of new derivatives by site-directed mutagenesis

The substrate specificities of the A domains was predicted by NRPSpredict2 and the putative active pockets were compared with specificity-conferring code [69, 70] (supplementary). Interestingly, the active residues of the A_1 domain (DVWHLSLI<u>E</u>K) differ from common A_{Ala} domains (DLLFGIAVLK) but are highly similar to A_{Ser} domains (DVWHLSLI<u>D</u>K), resulting in Ser as the predicted specificity. The only difference of A_1 to Ser activating A domains is found at positon 331, where a glutamic acid is found in place of an aspartic acid. This led us to the hypothesis, that an E331D mutation of the A_1 domain could change the substrate specificity to Ser. To test this hypothesis, the Glu_{331} encoding codon GAG was changed to GAC by site directed mutagenesis via *ccdB* counter selection. The mutant p15A-Ptet-vioArecov-A1E331D was transformed into *M. xanthus* DK1622 (*ΔmchA*) and fermented in CTT medium with supplement of pipecolic acid. HPLC-MS analysis showed that vioprolides were produced as hydroxylated forms (vioprolide B_1~D_1) (Fig. 2.6), eluting at earlier retention times compared to the vioprolides B and D. Notably, the mutated A_1 domain retained the capability to incorporate Ala, as the vioprolides B and D were still produced in low amounts. To validate the replacement of alanine by serine, we purified 1.4 mg vioprolide B_1 and 2.6 mg vioprolide D_1 from 10 L cultures and the structure of vioprolide D_1 was elucidated using NMR spectroscopy (supplementary). Surprisingly, the pronounced cytotoxicity of vioprolide B was lowered by a factor of 100 by hydroxylation. The fundamental reason of this observation remains hidden, as both the molecular target and the mode of action of vioprolide B are unknown at present. Nevertheless, the

introduction of a second hydroxyl group enables the generation of a larger variety of derivatives. Inspired by the engineering of the NRPS product pyochelin where the inactivation of an NMT domain by mutation of a conserved Gly into Arg led to the production of *N*-desmethyl derivatives [71], we tried to generate *N*-desmethyl vioprolides in a similar way. The NMT domain in module 10 was inactivated by mutating the conserved Gly encoding codon GGA to the Arg encoding codon CGA and the resulting mutant p15A-Ptet-vioArecov-NMTmut was transformed into *M. xanthus* DK1622 (*ΔmchA*). LC-MS analysis revealed the production of desmethyl-vioprolides (vioprolide B_2~D_2) (Fig. 2.6) at significantly lower levels than the original vioprolides. We therefore decided to limit the characterization to LC-MS and MS/MS. A simultaneous mutation of the A_1 domain and the NMT domain led to the production of hydroxyl-desmethyl-vioprolides (vioprolide B_3~D_3) in *M. xanthus* (Fig. 2.6) but the production was accompanied by a further drop in yield. As reported in pyochelin biosynthesis, inactivation of an NMT domain affected the chain release from the assembly line. The yield reduction of desmethyl-vioprolides might have resulted from stagnant chain release, which might be overcome by deletion of the NMT domain.
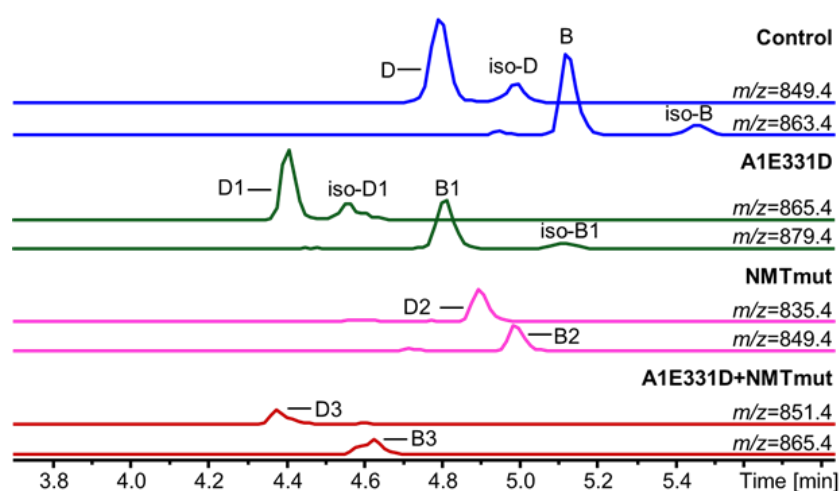

Fig. 2.6 UPLC-HRMS analysis of vioprolides and derivatives in *M. xanthus* mutants. B and D, vioprolide B and D; B1 and D1, vioprolide B1 and D1 from

A_1E331D mutants; B2 and D2, vioprolide B2 and D2 from NMT mutants; B3 and D3, vioprolides B3 and D3 from A_1E331D and NMT dual mutants; iso-B, iso-D, iso-B1 and iso-D1, isomers of the respective vioprolides. EIC *m/z* ([M+H]$^+$) of the respective compound is shown at the right side. Intensity is set to the same range.

Conclusion

Vioprolides were shown to be synthesized on an NRPS machinery with an uncommon initiation module that is responsible for the formation of acylated derivatives. We have shown that the FAAL domain activates long chain fatty acids and loads them onto the ACP domain. The FkbH domain transfers D-1, 3-bisphosphoglycerate onto PCP_1 domain and the unusual C_1 domain couples these substrates to deliver a 2-acyl-glyceric acid to the following NRPS machinery. The resulting acylated derivatives could only be found in the cell pellet which we believe to be attributed to a hydrolytic maturation process for facilitated compound export. In feeding studies we could prove that the carbon skeleton of methyl-azetidine is derived from proline, but no responsible protein could be identified in the cluster. The established heterologous expression system allowed an approx. 8-fold higher production level in half the time. Site directed mutagenesis of the A_1 domain of module 3 and the NMT domain in module 10 enabled us to generate analogues that, especially in the case of hydroxylation, broaden the scope of possible semi-synthetic derivatives. Our work sets the stage for future research to reveal the mode of action.

2.3 Experimental section

Bacterial strains, primers and culturing conditions

Strains and primers used in this study are shown in supplementary. *E. coli* GB2005 and HS996 were used for cloning, and *E. coli* GB05-Red was used for Red/ET recombination [72]. *E. coli* GBred-gyrA462 was used for site-directed mutagenesis [73]. *E. coli* Rosetta (DE3) and *E. coli* BL21 (DE3) were used to express proteins for *in vitro* assay. *E. coli* was cultivated at 37 ^{o}C in low salt Luria-Bertani (LB) broth (Trypton 10 g/L; Yeast extract 5 g/L; NaCl 1 g/L). *M. xanthus* DK1622 (*ΔmchA*) was propagated in CTT liquid medium (Casitone 10 g/L, 1 M Tris pH 7.6 10 ml/L, 1 M K_2HPO_4 1 ml/L, 0.8 M $MgSO_4$ 10 ml/L, pH 7.6) at 30 ^{o}C. *Burkholderia sp.* DSM7029 was cultivated in optimized CYCG medium (Casitone 6 g/L; Yeast extracts 2 g/L; Glycerol 0.5% v/v; $CaCl_2 \times 2H_2O$ 1.4 g/L). *P. putida* KT2440 was grown in LB medium and the conjugants were selected in *Pseudomonas* minimum medium [74]. Appropriate antibiotics were added when needed (ampicillin, 100 μg/mL; kanamycin, 50 μg/mL; gentamicin, 5 μg/mL; tetracycline, 5 μg/mL; oxytetracycline, 5 μg/mL; chloramphenicol, 20 μg/mL; blasticidin, 40 μg/mL; zeocin, 25 μg/mL). Plasmid transformation and Red/ET recombination conditions are given in supplementary.

Cloning and engineering of the *vio* gene cluster

The genome of *C. violaceus* Cb vi35 was sequenced by Illumina Mseq sequencing. SuperCos 1 plasmid was used to construct the cosmid genomic library according to the manufacturer's protocol. The vioprolide biosynthetic gene cluster (*vio*) was screened by colony hybridization. Several probes targeting on *vioZ*, *vioD*, *orf3* and *orf9* were generated by polymerase chain reaction (PCR) with the primer pairs ET21-cyclo-up / vio-5, Vio9-Cfor / Vio9-Crev, vio-13 / vio-14 and vio-9 / vio-10. The positive clones were verified

further by Illumina Mseq sequencing. The missing parts of the gene cluster were obtained by PCR and Red/ET recombination. The procedures for construction of *vio* expression vector, repairing of the mutation sites, gene knock-out and site-directed mutagenesis are shown in supplementary.

Bioinformatics analysis

The *vio* gene cluster was annotated with antiSMASH platform [8, 9]. Domains and related ORFs in *vio* gene cluster were aligned with NCBI non-redundant protein database by using the BLAST program (https://blast.ncbi.nlm.nih.gov). The specificity-conferring code of adenylation (A) domains was analyzed by NRPSpredictor2 web server (http://nrps.informatik.uni-tuebingen.de) and was compared with the selectivity-conferring code table of A domains [69].

Heterologous expression and analysis of vioprolides

M. xanthus mutants were fermented in 100 mL shaking flasks containing 25 mL CTT medium with kanamycin and oxytetracycline and were cultivated for 4.5 days at 30 °C on a rotary shaker (180rpm) , 0.7 mg/L pipecolic acid was added every 24 hours when necessary. *B. sp.* DSM7029 mutants were cultivated in 25 mL CYCG medium at 30 °C for 4.5 days, and *P. putida* mutants were fermented in 25 mL LB medium at 30 °C for 3 days. The absorber resin Amberlite XAD-16 (1%) was added 12 hours before compounds extraction. The XAD-16 resin was collected and compounds were extracted with methanol. The extracts were evaporated and dissolved in 1 mL methanol to make 25-fold concentrates, 1 μL concentrated extract was analyzed by high-performance liquid chromatography/electrospray ionization tandem mass spectrometry (HPLC-MS): a Thermo Scientific Dionex UltiMate 3000 Rapid Separation LC (RSLC) system coupled to amaZon speed ion trap mass spectrometer (Bruker Daltonics, Germany). Chromatographic separation was carried out on a Waters

Acquity UPLC BEH C18 column (50 × 2 mm, 1.7 μm particle size) with a solvent system consisting of ddH$_2$O + 0.1% formic acid (A) and acetonitrile + 0.1% formic acid (B) gradient at a flow rate of 0.6 mL/min and a temperature of 45 ^{o}C. Gradient: 0-0.5 min 5% B; 0.5-9.5 min linear from 5%-95% B; 9.5-10.5 min isocratic at 95% B; 10.5-11 min linear from 5% B. Vioprolides were identified by comparison to the retention time, molecular mass and MS/MS fragmentation patterns of the referential compounds (vioprolide A/B: *m/z* [M+H]$^+$=863.4; vioprolide C/D: *m/z* [M+H]$^+$=849.4).

Construction of expression plasmids for *in vitro* assay

The FAAL-ACP-C-FkbH-PCP (FAAL-PCP) encoding sequence *faal-pcp* was amplified from p15A-Ptet-vio with Phusion DNA polymerase (Fermentas) by using primers Module1-NheI-5 and Module1-NotI-3, and was then cloned to pET28b-sumo-tev at *Nhe* I and *Not* I sites to generate pET28b-STFP. The *faal-pcp* was also cloned to pET28b at *Nco* I and *Not* I sites to yield pET28b-FP which expresses FAAL-ACP with an N-terminal his-tag. The module 1 (FAAL-ACP) expression vector pET28b-STFA was constructed by deleting *c-fkbH-pcp* on the plasmid pET28b-STFP. The 3'-terminal sequence of *faal-acp* was PCR-amplified with primers FA-AatII-5 and FA-NotI3, and then replaced *c-fkbH-pcp* of the pET28b-STFP at *Aat* II and *Not* I sites. The encoding sequence of ACP domain (*acp*) was PCR-amplified with primers HisACP-NcoI5 and FA-NotI3, which was inserted into pET28b at *Nco* I and *Not* I sites to give pET28b-acp. The *fkbH-pcp* was amplified with primers KP-NheI5 and KP-NotI3, and was then cloned to pET28b-sumo-tev at *Nhe* I and *Not* I sites to generate pET28b-STKP. The final expression vectors were verified by restriction enzymes and sequencing.

***In vitro* protein activity assay**

Protein expression and purification procedures can be obtained in supplementary. Transformation of *apo*-FAAL-ACP and *apo*-ACP to their *holo*-forms was performed by incubating 5 μM protein with 1 μM MtaA (PPTase), 8 mM $MgCl_2$ and 1 mM CoA at 37 ^{o}C for 1h. The fatty acid loading activity of FAAL-ACP was initially performed in 50 μL reaction system with 5 μM FAAL-ACP, 1 μM MtaA (PPTase), 8 mM $MgCl_2$, 1 mM CoA, 100 μM fatty acid ($C_{16:0}$, *iso*-$C_{16:0}$, $C_{17:0}$ or *iso*-$C_{17:0}$) and 5 mM ATP. The reaction mixture was incubated at 37 ^{o}C for 1h and measured with LC-MS (MaXis 4G). The *in trans* loading fatty acids to ACP was performed in two steps: internal palmitic acid in FAAL domain was transferred to its downstream ACP by incubation with 5 μM *holo*-FAAL-ACP and 5 mM ATP at 30 ^{o}C for 30min. The final concentration of 10 μM *holo*-ACP, 100 μM fatty acid ($C_{16:0}$, *iso*-$C_{16:0}$, $C_{17:0}$ or *iso*-$C_{17:0}$) and 5 mM ATP were then added to the reaction system and incubated at 30 ^{o}C for another 30 min.

The offloading of FAAL-PCP attached intermediates followed the reported procedure [28]. The *apo*-FAAL-ACP was initially loaded with fatty acid and glyceric acid. The reaction was performed in 100 μL reaction mixture with 6.5 μM *apo*-FAAL-PCP, 1 μM MtaA, 0.625 U 3-phosphoglyceric phosphokinase (PGK, Sigma), 13 μM D-3-phosphoglyceric acid, 5 mM ATP, 100 μM fatty acid ($C_{17:0}$ or *iso*-$C_{17:0}$), 10 mM $MgCl_2$ and 1 mM CoA. The reaction system was incubated at 30 ^{o}C for 1h. Cysteamine was then added to a final concentration of 0.2 M and samples were stored at 4 °C overnight. The volume of all samples was adjusted to 500 μL with brine, prior to two rounds of extraction with ethylacetate. The combined ethylacetate phases were evaporated to dryness using an Eppendorf Concentrator plus (Hamburg, Germany). After reconstitution in 100 μL MeOH, samples were subjected to LC-FTICR-MS and –MS/MS characterization.

Measurement of intact proteins and protein-bound intermediates

ESI-MS-measurements of intact proteins were performed on a Dionex (Germering, Germany) Ultimate 3000 RSLC system using a ProSwift RP-4H (monolithic PS-DVB), 250 × 1 mm column (Thermo, USA). Separation of 1 μL sample was achieved by a multistep gradient from (A) H2O + 0.1% formic acid + 1% DMSO to (B) ACN + 0.1% formic acid + 1% DMSO at a flow rate of 200 μL/min and 45 °C. Chromatographic conditions were as follows: 0-0.5 min, 5% B; 0.5-18.5 min, 5-65% B; 18.5-19.5 min, 65-98% B; 19.5-21.5 min, 98% B; 21.5-23.5 min, 98-5% B; 23.5-26 min, 5% B. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μL/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker Apollo II ESI source. In the source region, the temperature was set to 180 °C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 L/min and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 600-1800 *m/z* at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

The offloaded intermediates from FAAL-PCP were measured on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 × 2.1 mm, 1.7 μm dp column (Waters, Germany). Separation of 1 μL sample was achieved by a linear gradient from (A) H2O + 0.1% formic acid to (B) ACN + 0.1% formic acid at a flow rate of 600 μL/min and 45 °C. Gradient conditions were as follows: 0-1 min, 5% B; 1-7 min, 5-95% B; 7-8.5 min, 95% B; 8.5-9 min, 95-5% B; 9-11.5 min, 5% B. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μL/min before entering the solariX XR (7T) FT-ICR mass spectrometer (Bruker Daltonics, Germany) using the Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4500 V, the dry-gas flow was 4.0 L/min and the nebulizer

was set to 1.1 bar. Mass spectra were acquired in positive ionization mode as reduced profile spectra ranging from 100-750 *m/z* with a FID data size of 1 M, a quadrupole accumulation time of 200 ms and data reduction set at 98%. Q-CID fragment spectra were recorded using a fragmentation voltage of 15 eV.

2.4 References

1 Schummer, D., Höfle, G., Forche, E., Reichenbach, H., Wray, V., Domke, T. (1996) Antibiotics from Gliding Bacteria, LXXVI. Vioprolides: New Antifungal and Cytotoxic Peptolides from *Cystobacter violaceus*. *Liebigs Annalen-Recueil*, **6**:971–978.

2 Pegorie, M., Denning, D.W., Welfare, W. (2016) Estimating the burden of invasive and serious fungal disease in the United Kingdom. *Journal of Infection* doi: 10.1016/j.jinf.2016.10.005.

3 Park, B.J., Chiller, T.M., Brandt, M.E., Warnock, D.W. Epidemiology of Systemic Fungal Diseases: An Overview, pp. 27–37.

4 Majer, O., Bourgeois, C., Zwolanek, F., Lassnig, C., Kerjaschki, D., Mack, M., Müller, M., Kuchler, K. (2012) Type I interferons promote fatal immunopathology by regulating inflammatory monocytes and neutrophils during *Candida* infections. *PLoS Pathogens*, **8** (7), e1002811.

5 Lorenzen, W., Bozhuyuk, K.A.J., Cortina, N.S., Bode, H.B. (2014) A comprehensive insight into the lipid composition of *Myxococcus xanthus* by UPLC-ESI-MS. *Journal of Lipid Research*, **55** (12), 2620–2633.

6 Garcia, R., Pistorius, D., Stadler, M., Müller, R. (2011) Fatty acid-related phylogeny of myxobacteria as an approach to discover polyunsaturated omega-3/6 Fatty acids. *Journal of Bacteriology*, **193** (8), 1930–1942.

7 Kearns, D.B., Venot, A., Bonner, P.J., Stevens, B., Boons, G.J., Shimkets, L.J. (2001) Identification of a developmental chemoattractant in *Myxococcus xanthus* through metabolic engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **98** (24), 13990–13994.

8 Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., Weber, T. (2013) antiSMASH 2.0 - a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, **41** (Web Server Issue), W204-W212.

9 Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E.,

Medema, M.H. (2015) antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, **43**(W1):W237-243.

10 Rottig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., Kohlbacher, O. (2011) NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, **39** (suppl_2), W362-W367.

11 Trivedi, O.A., Arora, P., Sridharan, V., Tickoo, R., Mohanty, D., Gokhale, R.S. (2004) Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature*, **428** (6981), 441–445.

12 Arora, P., Goyal, A., Natarajan, V.T., Rajakumara, E., Verma, P., Gupta, R., Yousuf, M., Trivedi, O.A., Mohanty, D., Tyagi, A., Sankaranarayanan, R., Gokhale, R.S. (2009) Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nature Chemical Biology*, **5** (3), 166–173.

13 Hayashi, T., Kitamura, Y., Funa, N., Ohnishi, Y., Horinouchi, S. (2011) Fatty Acyl-AMP Ligase Involvement in the Production of Alkylresorcylic Acid by a *Myxococcus xanthus* Type III Polyketide Synthase. *ChemBioChem*, **12** (14), 2166–2176.

14 Reimer, D., Pos, K.M., Thines, M., Grün, P., Bode, H.B. (2011) A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nature Chemical Biology*, **7** (12), 888–890.

15 Zhu, Y., Zhang, W., Chen, Y., Yuan, C., Zhang, H., Zhang, G., Ma, L., Zhang, Q., Tian, X., Zhang, S., Zhang, C. (2015) Characterization of Heronamide Biosynthesis Reveals a Tailoring Hydroxylase and Indicates Migrated Double Bonds. *ChemBioChem*, **16** (14), 2086–2093.

16 Müller, S., Garcia-Gonzalez, E., Mainz, A., Hertlein, G., Heid, N.C., Mosker, E., van den Elst, H., Overkleeft, H.S., Genersch, E., Sussmuth, R.D. (2014) Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*. *Angewandte Chemie International Edition*, **53** (40), 10821–10825.

17 Schalk, I.J. and Guillon, L. (2013) Pyoverdine biosynthesis and secretion in *Pseudomonas aeruginosa*: implications for metal homeostasis. *Environmental Microbiology*, **15** (6), 1661–1673.

18 Brotherton, C.A. and Balskus, E.P. (2013) A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *Journal of the American Chemical Society*, **135** (9), 3359–3362.

19 Pu, J.-Y., Peng, C., Tang, M.-C., Zhang, Y., Guo, J.-P., Song, L.-Q., Hua, Q., Tang, G.-L. (2013) Naphthyridinomycin biosynthesis revealing the use of

leader peptide to guide nonribosomal peptide assembly. *Organic Letters*, **15** (14), 3674–3677.

20 Hiratsuka, T., Koketsu, K., Minami, A., Kaneko, S., Yamazaki, C., Watanabe, K., Oguri, H., Oikawa, H. (2013) Core assembly mechanism of quinocarcin/SF-1739: bimodular complex nonribosomal peptide synthetases for sequential mannich-type reactions. *Chemistry & Biology*, **20** (12), 1523–1535.

21 Fu, C., Keller, L., Bauer, A., Brönstrup, M., Froidbise, A., Hammann, P., Herrmann, J., Mondesert, G., Kurz, M., Schiell, M., Schummer, D., Toti, L., Wink, J., Müller, R. (2015) Biosynthetic Studies of Telomycin Reveal New Lipopeptides with Enhanced Activity. *Journal of the American Chemical Society*, **137** (24), 7692–7705.

22 Xu, Y., Kersten, R.D., Nam, S.-J., Lu, L., Al-Suwailem, A.M., Zheng, H., Fenical, W., Dorrestein, P.C., Moore, B.S., Qian, P.-Y. (2012) Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *Journal of the American Chemical Society*, **134** (20), 8625–8632.

23 Koketsu, K., Watanabe, K., Suda, H., Oguri, H., Oikawa, H. (2010) Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms. *Nature Chemical Biology*, **6** (6), 408–410.

24 Davidsen, J.M., Bartley, D.M., Townsend, C.A. (2013) Non-ribosomal propeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI. *Journal of the American Chemical Society*, **135** (5), 1749–1759.

25 Reimer, D. and Bode, H.B. (2014) A natural prodrug activation mechanism in the biosynthesis of nonribosomal peptides. *Natural Product Reports*, **31** (2), 154–159.

26 Trautman, E. and Crawford, J. (2016) Linking Biosynthetic Gene Clusters to their Metabolites via Pathway- Targeted Molecular Networking. *Current Trends in Medicinal Chemistry*, **16** (15), 1705–1716.

27 Cronan, J.E. and Thomas, J. (2009) Chapter 17 Bacterial Fatty Acid Synthesis and its Relationships with Polyketide Synthetic Pathways, in *Polyketides, aminocoumarins and carbohydrates* (ed D.A. Hopwood), Elsevier, Acad. Press, Amsterdam [u.a.], pp. 395–433.

28 Belecki, K. and Townsend, C.A. (2013) Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *Journal of the American Chemical Society*, **135** (38), 14339–14348.

29 Wu, K., Chung, L., Revill, W.P., Katz, L., Reeves, C.D. (2000) The FK520 gene cluster of *Streptomyces hygroscopicus var. ascomyceticus* (ATCC

14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene*, **251** (1), 81–90.

30 Emmert, E.A.B., Klimowicz, A.K., Thomas, M.G., Handelsman, J. (2004) Genetics of Zwittermicin A production by *Bacillus cereus*. *Applied and Environmental Microbiology*, **70** (1), 104–113.

31 Dorrestein, P.C., van Lanen, S.G., Li, W., Zhao, C., Deng, Z., Shen, B., Kelleher, N.L. (2006) The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. *Journal of the American Chemical Society*, **128** (32), 10386–10387.

32 Sun, Y., Hong, H., Gillies, F., Spencer, J.B., Leadlay, P.F. (2008) Glyceryl-S-acyl carrier protein as an intermediate in the biosynthesis of tetronate antibiotics. *ChemBioChem: a European journal of chemical biology*, **9** (1), 150–156.

33 Li, W., Ju, J., Rajski, S.R., Osada, H., Shen, B. (2008) Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces spiroverticillatus* unveiling new insights into dialkylmaleic anhydride and polyketide biosynthesis. *The Journal of Biological Chemistry*, **283** (42), 28607–28617.

34 Sun, Y., Hahn, F., Demydchuk, Y., Chettle, J., Tosin, M., Osada, H., Leadlay, P.F. (2010) *In vitro* reconstruction of tetronate RK-682 biosynthesis. *Nature Chemical Biology*, **6** (2), 99–101.

35 Zhang, F., He, H.Y., Tang, M.C., Zhou, Q., Tang, G.L. (2011) Cloning and elucidation of the FR901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate as starter units. *Journal of the American Chemical Society*, **138** (8), 2452–2462.

36 Jahns, C., Hoffmann, T., Müller, S., Gerth, K., Washausen, P., Höfle, G., Reichenbach, H., Kalesse, M., Müller, R. (2012) Pellasoren: structure elucidation, biosynthesis, and total synthesis of a cytotoxic secondary metabolite from *Sorangium cellulosum*. *Angewandte Chemie (International ed. in English)*, **51** (21), 5239–5243.

37 Etzbach, L., Plaza, A., Garcia, R., Baumann, S., Müller, R. (2014) Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Organic Letters*, **16** (9), 2414–2417.

38 Bertin, M.J., Vulpanovici, A., Monroe, E.A., Korobeynikov, A., Sherman, D.H., Gerwick, L., Gerwick, W.H. (2016) The Phormidolide Biosynthetic

Gene Cluster: A *trans*-AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide. *ChemBioChem*, **17** (2), 164–173.

39 Chen, C.-Y., Han, W.-B., Chen, H.-J., Wu, Y., Gao, P. (2013) Optically Active Monoacylglycerols: Synthesis and Assessment of Purity. *European Journal of Organic Chemistry*, **2013** (20), 4311–4318.

40 Chen, H.-J., Chen, C.-Y., Gao, P., Wu, Y. (2013) On the optical rotation of 1 (or 3)-stearoyl-sn-glycerol. *Tetrahedron*, **69** (46), 9848–9851.

41 Rosseto, R. and Hajdu, J. (2014) Synthesis of phospholipids on a glyceric acid scaffold: design and preparation of phospholipase A2 specific substrates. *Tetrahedron*, **70** (19), 3155–3165.

42 Rosseto, R., Tcacenco, C.M., Ranganathan, R., Hajdu, J. (2008) Synthesis of Phosphatidylcholine Analogues Derived from Glyceric Acid: a New Class of Biologically Active Phospholipid Compounds. *Tetrahedron Letters*, **49** (21), 3500–3503.

43 Julien, B. and Shah, S. (2002) Heterologous Expression of Epothilone Biosynthetic Genes in *Myxococcus xanthus*. *Antimicrobial Agents and Chemotherapy*, **46** (9), 2772–2778.

44 Stevens, D.C., Henry, M.R., Murphy, K.A., Boddy, C.N. (2010) Heterologous expression of the oxytetracycline biosynthetic pathway in *Myxococcus xanthus*. *Applied and Environmental Microbiology*, **76** (8), 2681–2683.

45 Fu, J., Wenzel, S.C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A.F., Müller, R., Zhang, Y. (2008) Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Research*, **36** (17), e113.

46 Tu, Q., Herrmann, J., Hu, S., Raju, R., Bian, X., Zhang, Y., Müller, R. (2016) Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. *Scientific Reports*, **6**, 21066.

47 Tang, L., Chung, L., Carney, J.R., Starks, C.M., Licari, P., Katz, L. (2005) Generation of new epothilones by genetic engineering of a polyketide synthase in *Myxococcus xanthus*. *The Journal of Antibiotics*, **58** (3), 178–184.

48 Chai, Y., Shan, S., Weissman, K.J., Hu, S., Zhang, Y., Müller, R. (2012) Heterologous expression and genetic engineering of the tubulysin biosynthetic gene cluster using Red/ET recombineering and inactivation mutagenesis. *Chemistry & Biology*, **19** (3), 361–371.

49 Perlova, O., Fu, J., Kuhlmann, S., Krug, D., Stewart, A.F., Zhang, Y., Müller, R. (2006) Reconstitution of the myxothiazol biosynthetic gene cluster by

Red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Applied and Environmental Microbiology*, **72** (12), 7485–7494.

50 Liu, X. and Cheng, Y.-Q. (2014) Genome-guided discovery of diverse natural products from *Burkholderia sp.*. *Journal of Industrial Microbiology & Biotechnology*, **41** (2), 275–284.

51 Esmaeel, Q., Pupin, M., Kieu, N.P., Chataigne, G., Bechet, M., Deravel, J., Krier, F., Hofte, M., Jacques, P., Leclere, V. (2016) *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. *MicrobiologyOpen*, **5** (3), 512–526.

52 Depoorter, E., Bull, M.J., Peeters, C., Coenye, T., Vandamme, P., Mahenthiralingam, E. (2016) *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. *Applied Microbiology and Biotechnology*, **100** (12), 5215–5229.

53 Schellenberg, B., Bigler, L., Dudler, R. (2007) Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. *Environmental Microbiology*, **9** (7), 1640–1650.

54 Bian, X., Huang, F., Wang, H., Klefisch, T., Müller, R., Zhang, Y. (2014) Heterologous production of glidobactins/luminmycins in *Escherichia coli* nissle containing the glidobactin biosynthetic gene cluster from *Burkholderia* DSM7029. *ChemBioChem*, **15** (15), 2221–2224.

55 Loeschcke, A. and Thies, S. (2015) *Pseudomonas putida*-a versatile host for the production of natural products. *Applied Microbiology and Biotechnology*, **99** (15), 6197–6214.

56 Ongley, S.E., Bian, X., Zhang, Y., Chau, R., Gerwick, W.H., Müller, R., Neilan, B.A. (2013) High-titer heterologous production in *E. coli* of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. *ACS Chemical Biology*, **8** (9), 1888–1893.

57 Kawai, S., Itoh, K., Takagi, S.-i., Iwashita, T., Nomoto, K. (1988) Studies on phytosiderophores: Biosynthesis of mugineic acid and 2′-deoxymugineic acid in *Hordeum vulgare L. var. Minorimugi*. *Tetrahedron Letters*, **29** (9), 1053–1056.

58 Di, Y.-T., He, H.-P., Wang, Y.-S., Li, L.-B., Lu, Y., Gong, J.-B., Fang, X., Kong, N.-C., Li, S.-L., Zhu, H.-J., Hao, X.-J. (2007) Isolation, X-ray crystallography, and computational studies of calydaphninone, a new alkaloid from *Daphniphyllum calycillum*. *Organic Letters*, **9** (7), 1355–1358.

59 Kitajima, M., Kogure, N., Yamaguchi, K., Takayama, H., Aimi, N. (2003) Structure reinvestigation of gelsemoxonine, a constituent of Gelsemium

elegans, reveals a novel, azetidine-containing indole alkaloid. *Organic Letters*, **5** (12), 2075–2078.

60 Leete, E., Louters, L.L., Prakash Rao, H.S. (1986) Biosynthesis of azetidine-2-carboxylic acid in *Convallaria majalis*: Studies with N-15 labelled precursors. *Phytochemistry*, **25** (12), 2753–2758.

61 Kobayashi, J., Cheng, J.-F., Ishibashi, M., Wälchli, M.R., Yamamura, S., Ohizumi, Y. (1991) Penaresidin A and B, two novel azetidine alkaloids with potent actomyosin ATPase-activating activity from the Okinawan marine sponge *Penares sp.. Journal of the Chemical Society, Perkin Transactions 1* 1 (5), 1135–1137.

62 Hayashi, H., Takiuchi, K., Murao, S., Arai, M. (1988) Okaramine B, an insecticidal indole alkaloid, produced by *Penicillium simplicissimum* AK-40. *Agricultural and Biological Chemistry*, **52** (8), 2131–2133.

63 Kagiyama, I., Kato, H., Nehira, T., Frisvad, J.C., Sherman, D.H., Williams, R.M., Tsukamoto, S. (2016) Taichunamides: Prenylated Indole Alkaloids from *Aspergillus taichungensis* (IBT 19404). *Angewandte Chemie (International ed. in English)*, **55** (3), 1128–1132.

64 Funayama, S. and Isono, K. (1975) Biosynthesis of the polyoxins, nucleoside peptide antibiotics. Glutamate as an origin of 2-amino-2-deoxy-L-xylonic acid (polyoxamic acid). *Biochemistry*, **14** (26), 5568–5572.

65 Chen, W., Huang, T., He, X., Meng, Q., You, D., Bai, L., Li, J., Wu, M., Li, R., Xie, Z., Zhou, H., Zhou, X., Tan, H., Deng, Z. (2009) Characterization of the polyoxin biosynthetic gene cluster from *Streptomyces cacaoi* and engineered production of polyoxin H. *Journal of Biological Chemistry*, **284** (16), 10627–10638.

66 Hill, R.K., Rhee, S.W., Isono, K., Crout, D.H.G., Suhadolnik, R.J. (1981) Stereospecificity of the enzymic dehydrogenation in the biosynthesis of 3-ethylidene-L-azetidine-2-carboxylic acid from isoleucine by *Streptomyces cacaoi*. *Biochemistry*, **20** (24), 7040–7042.

67 Bachmann B.O., Li R., Townsend C.A. (1998) beta-Lactam synthetase: a new biosynthetic enzyme. *Proceedings of the National Academy of Sciences*, **95**(16):9082-9086.

68 Gaudelli N.M., Long D.H., Townsend C.A. (2015) β-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature*, **520**(7547):383-387.

69 Stachelhaus, T., Mootz, H.D., Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, **6**, 493–505.

70 Challis, G.L., Ravel, J., Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology*, **7** (3), 211–224.

71 Patel, H.M. and Walsh, C.T. (2001) In Vitro Reconstitution of the *Pseudomonas aeruginosa* Nonribosomal Peptide Synthesis of Pyochelin: Characterization of Backbone Tailoring Thiazoline Reductase and N - Methyltransferase Activities †. *Biochemistry*, **40** (30), 9023–9031.

72 Zhang, Y., Muyrers, J.P.P., Testa, G., Stewart, A.F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nature Biotechnology*, **18**, 1314–1317.

73 Wang, H., Bian, X., Xia, L., Ding, X., Müller, R., Zhang, Y., Fu, J., Stewart, A.F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection. *Nucleic Acids Research*, **42** (5), e37.

74 Wenzel, S.C., Gross, F., Zhang, Y., Fu, J., Stewart, F.A., Müller, R. (2005) Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chemistry & Biology*, **12** (3), 349–356.

2.5 Supplementary

Identification of previoprolides

  C. violaceus Cb vi35 was cultivated in 20 mL fresh Cas medium (Bacto-Casiton 10 g/L, $MgSO_4 \times 7H_2O$ 1 g/L, HEPES 10 g/L) in 100 mL flask at 30 °C for 3 days. Six milligrams isotope (L-Methionine-d3, L-Methionine-^{13}C, L-Methionine-^{13}C$_5$^{15}N, L-Threonine-^{13}C$_4$^{15}N or L-Proline-^{13}C$_5$^{15}N) was added in four portions to the culture at the time point of 0, 72, 96, 120 h after inoculation. *C. violaceus* Cb vi35 fermented without feeding was set as negative control. After fermentation XAD-16 absorber resin (1%) was added and incubated with the culture overnight. Compounds were extracted with methanol (1:1), evaporated and re-dissolved in 500 μL methanol, and 1 μL crude extract was analyzed by UPLC-HRMS.

Fig. S1 Identification of previoprolides from *C. violaceus* Cb vi35. UPLC-HRMS profiles of the compounds from cell medium (supernatant) and cell pellet are shown separately. EIC *m/z* of vioprolides (green) and previoprolides (red) are shown. Intensity is adjusted to the same range.
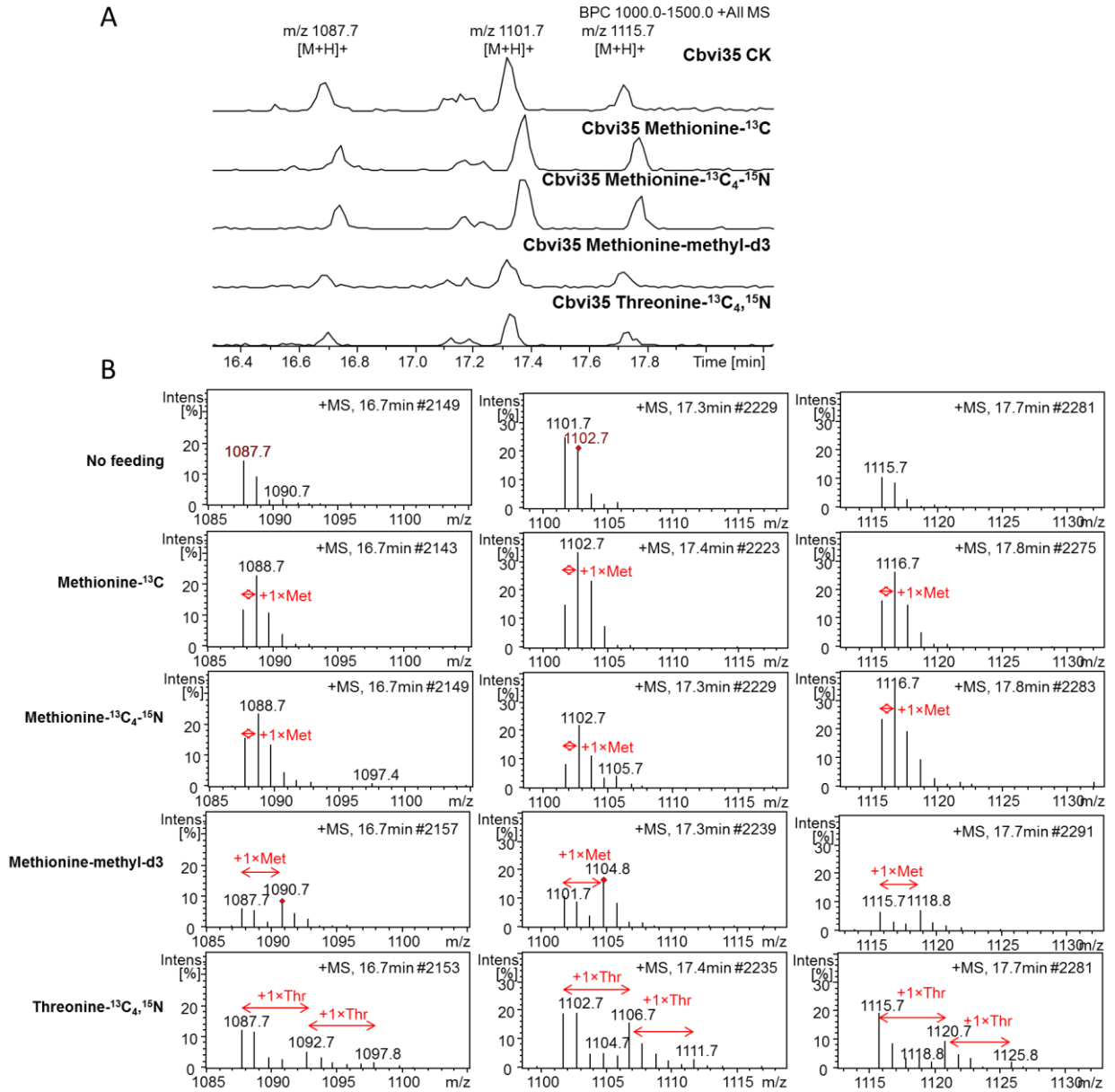
Fig. S2 HPLC-MS analysis of isotopically labeled previoprolides in *C. violaceus* Cb vi35. (A) HPLC profile of the previoprolides, the base peak (BPC) *m/z* 1000-1500 is shown; (B) mass spectra of previoprolides. The number of labeled methionine and theronine is shown in red arrows.
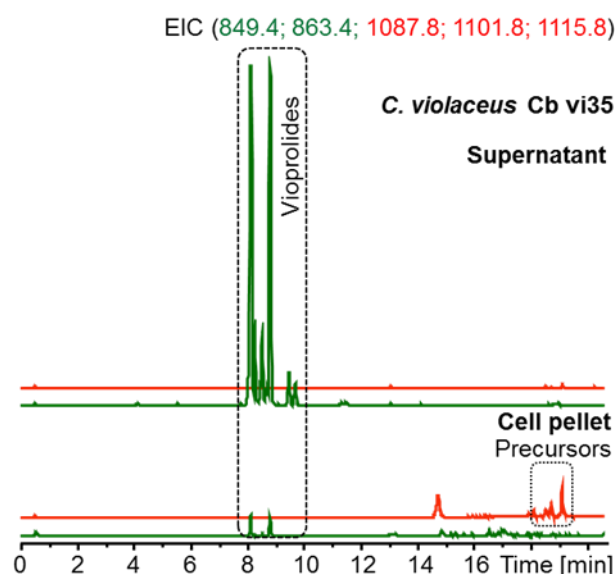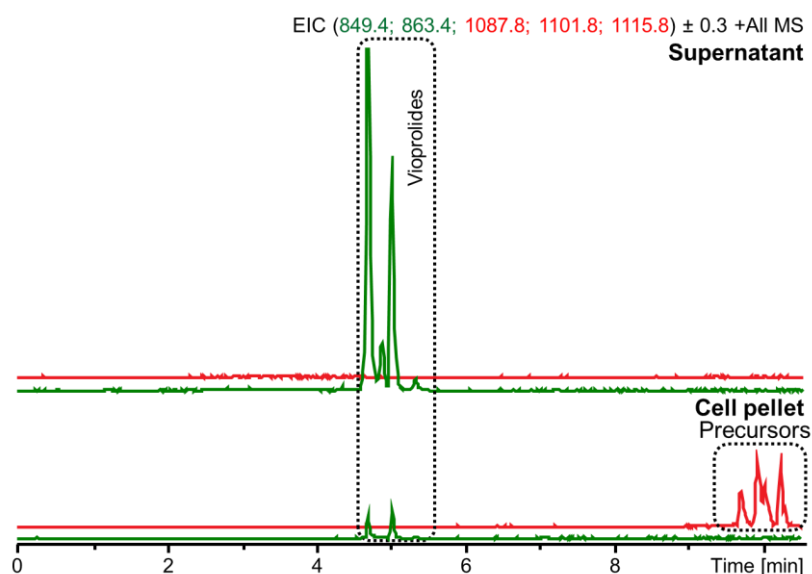
Fig. S3 Identification of previoprolides in *M. xanthus*::Ptet-vio. UPLC-HRMS profiles of the compounds from cell medium (supernatant) and cell pellet are shown separately. EIC *m/z* of vioprolides (green) and previoprolides (red) are shown. Intensity is adjusted to the same range.

Fig. S4 Identification of previoprolide in *Burkholderia sp.*DSM7029::Ptet-vio. HPLC profiles of the compounds from cell medium (supernatant) and cell pellet are shown separately. Most of the vioprolide B (indicated with dashed box) was secreted into cell medium, whereas previoprolide was intercepted in bacterial cells. EIC *m/z* of vioprolides (green) and previoprolides (red) are shown. Intensity is adjusted to the same range.
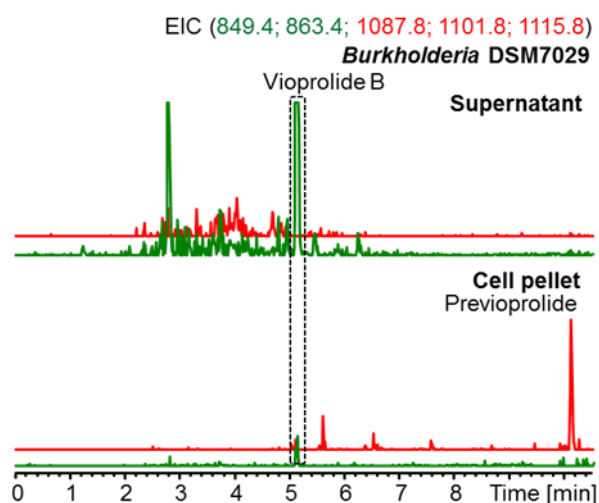
Characterization of vioprolide biosynthetic gene cluster

Table S1 Proposed function of the genes in the scaffold000098

Gene	Size (aa)	Proposed Function *	Identity	Accession No.
orf2	328	Hypothetical protein	36%	CRM73716.1
orf1	437	Transposase	58%	WP_014399689.1
vioZ	378	Ornithine cyclodeaminase	60%	WP_013376531.1
vioA	4913	NRPS: FAAL (23-570), ACP (596-677), C (705-1001), FkbH (1363-1680), PCP (1782-1848), E (1865-2167), C (2336-2634), A (2804-3284), PCP (3298-3369), C (3392-3690), A (3868-4342), PCP (4363-4423), E (4441-4742)	44%	WP_050045606.1
vioB	2159	NRPS: C (4-306), A (478-974), PCP (998-1061), Cy (1098-1396), A (1573-2053), PCP (2075-2139)	45%	WP_006635040.1
vioC	3320	NRPS: C (85-381), A (556-1043), PCP (1068-1129), C (1160-1459), A (1627-2113), PCP (2136-2200), C (2221-2526), A (2699-3191), PCP (3214-3275)	50%	KYC42747.1
vioD	1795	NRPS: C (43-343), A (515-951), NMT (1010-1131), PCP (1439-1502), TE (1524-1778)	46%	WP_002733092.1
orf3	600	Unusual protein kinase	32%	WP_006512948.1
orf4	443	Hypothetical protein Y590_21885	35%	AMB47606.1
orf5	550	Biotin synthase 1	56%	KKS93718.1
orf6	237	SAM-dependent methyltransferase	31%	WP_066486480.1
orf7	96	Hypothetical protein	40%	WP_040137362.1
orf8	96	Hypothetical protein	38%	WP_024750041.1
orf9	70	Antibiotic synthesis protein MbtH	89%	WP_043389803.1
orf10	362	GphU	98%	AHA38214.1
orf11	118	Response regulator	99%	WP_043408217.1
orf12	546	Molecular chaperone GroEL	99%	WP_043408220.1
orf13	518	Chemotaxis protein CheY	97%	WP_043408223.1
orf14	131	Ferritin	100%	WP_043408226.1
orf15	458	Diguanylate cyclase response regulator	99%	WP_043408229.1
orf16	257	Hypothetical protein	98%	WP_043408232.1

* Positions of NRPS domains are indicated in brackets; FAAL, fatty acyl-AMP ligase; ACP, acyl-carrier-protein domain; C, condensation domain; PCP, peptidyl-carrier-protein domain; E, epimerase; A, adenylation domain; Cy, heterocyclization domain; NMT, *N*-methyltransferase; TE, thioesterase.

Table S2 Specificity-conferring codes of the A domains in vioprolide NRPSs

A domains	Active residues (position)									
	235	236	239	278	299	301	322	330	331	517
A1: Ala	D	V	W	H	L	S	L	I	E	K
A2: Leu	D	A	W	F	L	G	H	V	V	K
A3: AZE/Pro	D	V	Q	C	L	S	E	V	T	K
A4: Cys	D	L	Y	N	L	S	L	I	W	K
A5: Dhb (Thr)	D	F	W	N	I	G	M	V	H	K
A6: Pip/Pro	D	I	Q	Y	Y	A	Q	V	V	K
A7: Thr	D	F	W	N	I	G	M	V	H	K
A8: Val	D	A	F	W	L	G	G	T	F	K

Sequence analysis of C domains

C domain	Motif C1	Motif C2	Motif C3
Vio-C$_1$	LPLSRHQLRLWWLE	ARYQLHLALRLEGDLQAHLLVRALELLVERHEALRATFS	VHHLVADGLSLEILVRELVASY
Vio-C$_2$	YPLAPVQQGMLFHT	GVYLEQWRARLRGPLDARALRAALEGIIERHASLRTAFQ	LHHLLLDAWSGLLVFREFITLY
Vio-C$_3$	PPVSFAQQRLWFLH	HAYHVVHALRLSGRLDTGAFERGLEEVVRRHETLRTTFR	MHHIISDGWSMGVLVRELAELY
Vio-C$_4$	YPLTPMQQGMLYHS	EVYFEQMCCSLRGRLVPAVLERAWQEVVNRHPALRTAFL	HHHLLLDGWSLPRLLKEVFTCY
Vio-C$_6$	LAASFAQQRLWLVD	PAYNLAFGIRMTGPVDAVALEEALNECVRRHEALRTRFT	MHHIVSDGWSLGVLAWELLTLY
Vio-C$_7$	YPLSMLQGGMVFHG	AIYHDIFSFHVRAPFEPEAWRGAVSELLRHHPVLRTSIE	FHHVILDGWSNAMMLTELNQRY
Vio-C$_8$	LPLSFGQQRLWFLE	AAYHMFVPLRADGVLEPVLLKRCLEALVRRHESLRTTFV	LHHIIADGWSLGVLVRELAALY
Vio-C$_9$	VPLSFAQRRLWFLH	AAYHMTYALRLSGELRVGALEASFAAIVQRHEALRTGFS	MHHIVSDGWSMGVIVRELVALY

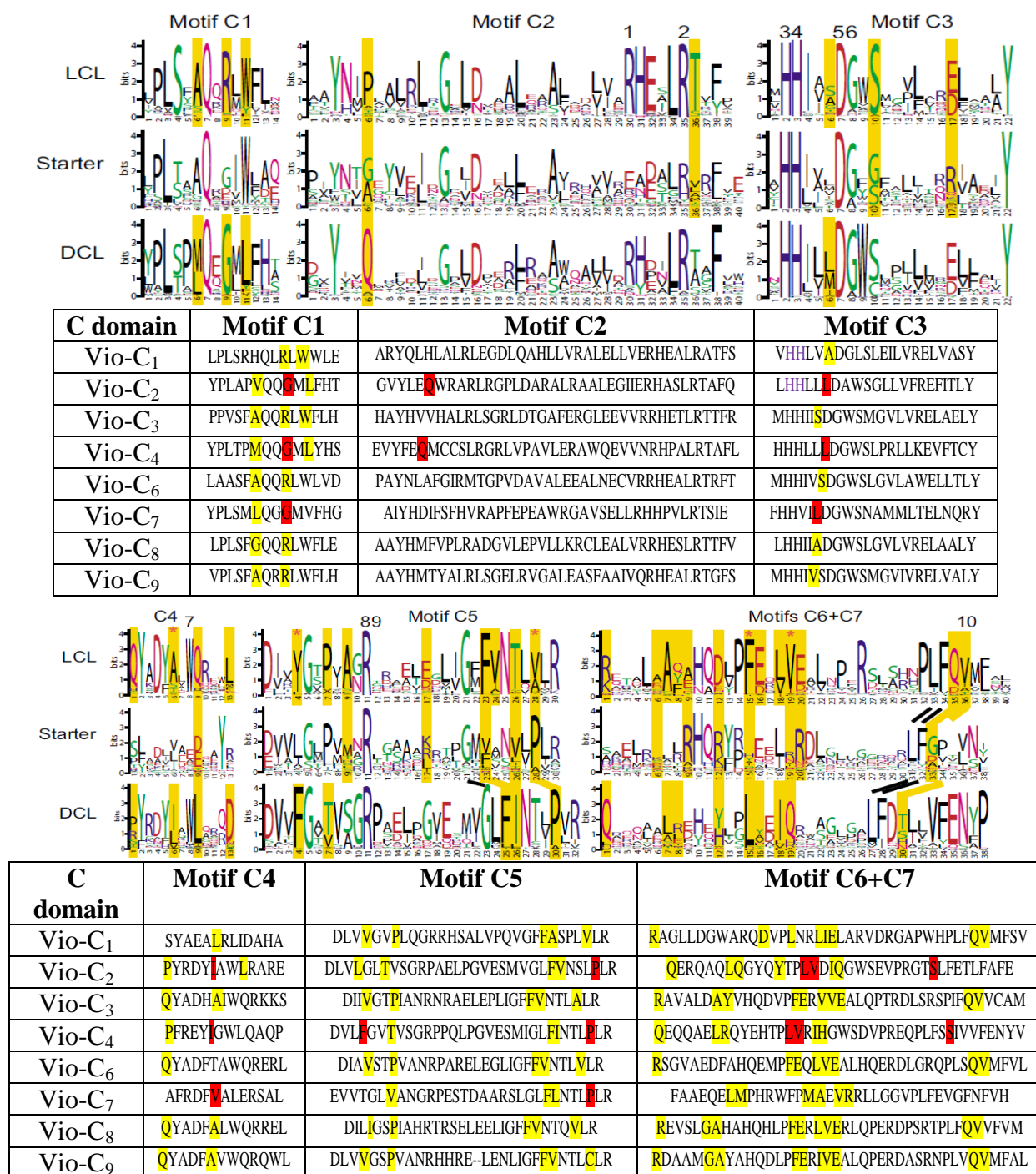| C domain	Motif C4	Motif C5	Motif C6+C7
Vio-C$_1$	SYAEALRLIDAHA	DLVVGVPLQGRRHSALVPQVGFFASPLVLR	RAGLLDGWARQDVPLNRLIELARVDRGAPWHPLFQVMFSV
Vio-C$_2$	PYRDYIAWLRARE	DLVLGLTVSGRPAELPGVESMVGLFVNSLPLR	QERQAQLQGYQYTPLVDIQGWSEVPRGTSLFETLFAFE
Vio-C$_3$	QYADHAIWQRKKS	DIIVGTPIANRNRAELEPLIGFFVNTLALR	RAVALDAYVHQDVPFERVVEALQPTRDLSRSPIFQVVCAM
Vio-C$_4$	PFREYIGWLQAQP	DVLSGVTVSGRPPQLPGVESMIGLFINTLPLR	QEQQAELRQYEHTPLVRIHGWSDVPREQPLFSSIVVFENYV
Vio-C$_6$	QYADFTAWQRERL	DIAVSTPVANRPARELEGLIGFFVNTLVLR	RSGVAEDFAHQEMPFEQLVEALHQERDLGRQPLSQVMFVL
Vio-C$_7$	AFRDFVALERSAL	EVVTGLVANGRPESTDAARSLGLFLNTLPLR	FAAEQELMPHRWFPMAEVRRLLGGVPLFEVGFNFVH
Vio-C$_8$	QYADFALWQRREL	DILIGSPIAHRTRSELEELIGFFVNTQVLR	REVSLGAHAHQHLPFERLVERLQPERDPSRTPLFQVVFVM
Vio-C$_9$	QYADFAVWQRQWL	DLVVGSPVANRHHRE--LENLIGFFVNTLCLR	RDAAMGAYAHQDLPFERIVEALQPERDASRNPLVQVMFAL

Fig. S5 Core motif analyses of C domains in vioprolide NRPSs. The crucial residues of $^{D}C_L$ domains are illustrated in red, and other characteristic residues are labeled in yellow. Comparing to the conserved residues reported by Rausch [1], C$_2$, C$_4$ and C$_7$ seem to be $^{D}C_L$ domain while others are $^{L}C_L$ domain.
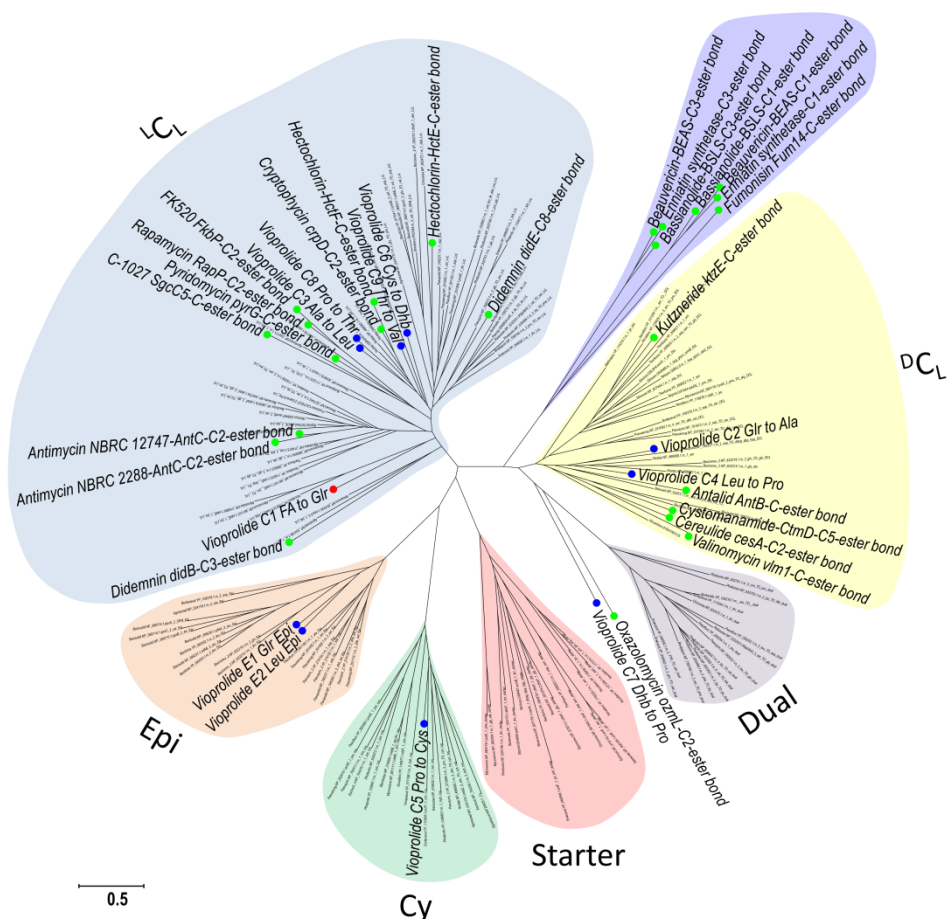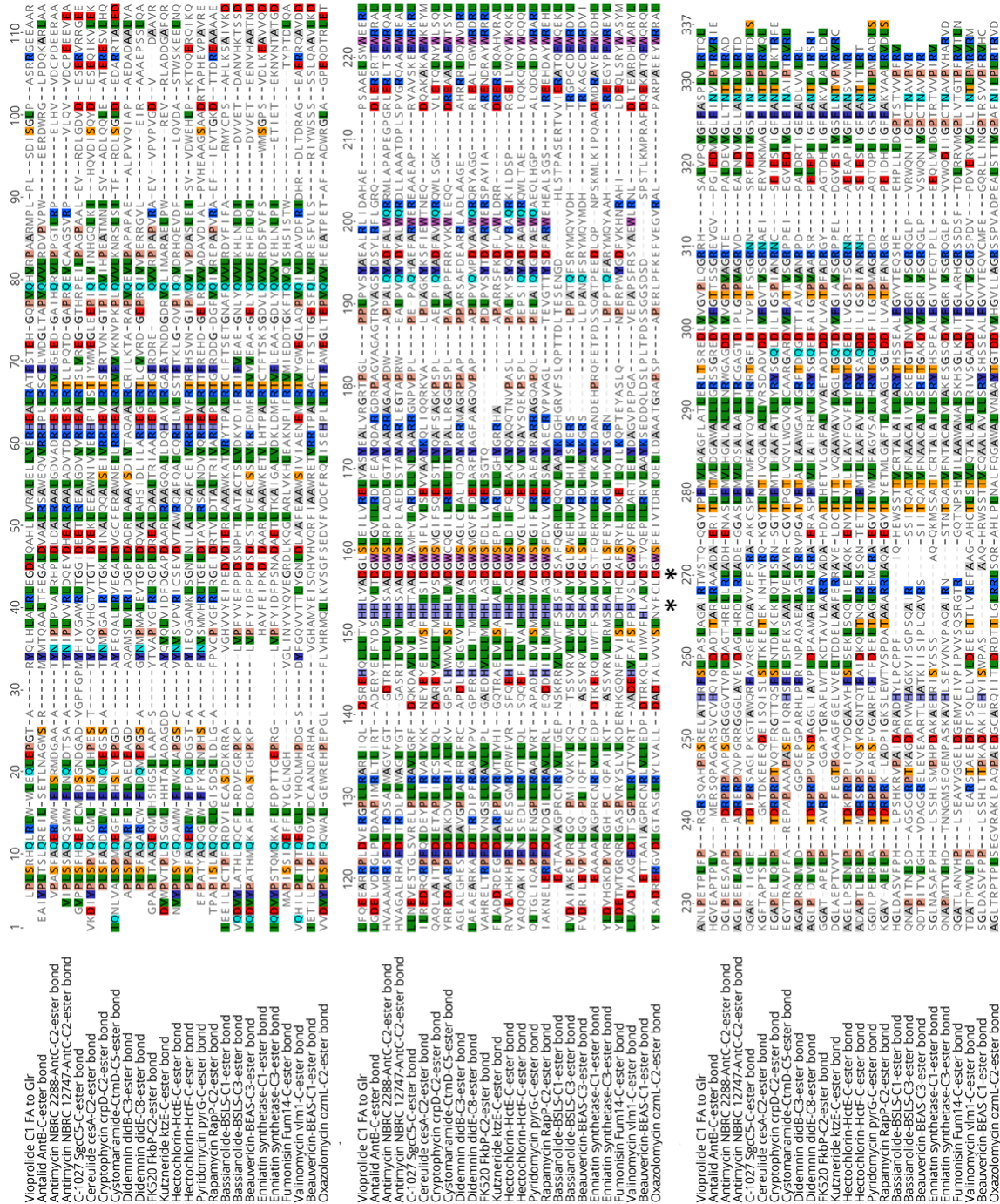
79

Fig. S6 Phylogenetic analysis of ester bond forming C domains and the C domains in vioprolide NRPSs. Phylogenetic tree was generated by FastTree plugin in Geneious R9. The C, E and Cy domains in vioprolide NRPS are marked with blue dots, vioprolide C_1 domain is labled with red dot and other ester bond forming C domains are indicated with green dots. ^LC_L, C domain condensates L-donor and L-acceptor; ^DC_L, C domain condensates D-donor and L-acceptor; Dual, C domain with both condensation and epimerization functions; Cy, cyclization domain; Epi, epimerization domain; Starter, starting C domain. Evolutionary distance is shown in scale bar.

Fig. S7 Sequence alignment of ester bond forming C domains (EBFCs). In total of 24 EBFC sequences were aligned with MUSCLE, the conserved His and Asp in the catalytic motif are labeled with asterisks.

In vitro experiments

The expression vector pET28b-STFP, pET28b-FP and pET28b-STFA (Fig. S8) were transformed into *E. coli* Rosetta (DE3), and pET28b-ACP and pET28b-STKP were transformed into *E. coli* BL21 (DE3). The transformants were selected on LB-agar plates with kanamycin. The recombinant protein expressing strains were cultivated in LB liquid medium at 37 $^{\circ}$C overnight. The overnight culture (1% v/v) of *E. coli* Rosetta (DE3)::pET28b-STFP, *E. coli* Rosetta (DE3)::pET28b-FP, *E. coli* BL21 (DE3)::pET28b-ACP or *E. coli* BL21 (DE3)::pET28b-STKP was then inoculated to 4×1.5 L LB. IPTG in an final concentration of 0.1 mM was added when OD_{600} reached up to 0.6. Cells were grown at 16 $^{\circ}$C for an additional 20 hours before collection. The *E. coli* Rosetta (DE3)::pET28b-STFA was inoculated to 4×1.5 L ZYM5052 auto-induction medium and cultivated at 37 $^{\circ}$C for 4 hours before growing at 16 $^{\circ}$C for 20 hours. The overnight culture (2% v/v) of MtaA expressing strain BL21 (DE3)::pCold-MtaA+pGro7 (a gift from Dr. Hilda Sucipto) was inoculated into 1 L LB broth with supplement of 2 mg/mL L-Arabinose, cells were cultivated at 37 $^{\circ}$C until OD_{600} reached up to 0.6. The cells were induced with 0.1 mM IPTG and allowed to grow at 16 $^{\circ}$C overnight.

Cells were harvested by centrifugation (8000 rpm, 10min, 4 $^{\circ}$C) and suspended in 100 mL washing buffer ([50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0] for FAAL-PCP, FAAL-ACP, ACP and FkbH; [20 mM Bis-Tris pH 6.8, 200 mM NaCl, 20 mM imidazole, 10% glycerol] for MtaA), and then disrupted by M-110P Microfluidizer® (Microfluidics). A proteinase inhibitor cocktail tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Roch) and 1 mg *DNAse* I (Roch) were added before cell disruption. Lysate was clarified by high speed centrifugation (20000 rpm, 30 min, 4 $^{\circ}$C) and the supernatant was loaded to HisTrap HP 5 mL column (GE Healthcare) connected with the HPLC system (ÄKTA avant 25, GE healthcare). After

washing with 10 CV (column volume) washing buffer, the recombinant protein was eluted with gradient elution buffer ([50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0] for FAAL-PCP, FAAL-ACP, ACP and FkbH; [20 mM Bis-Tris pH 6.8, 200 mM NaCl, 250 mM imidazole, 10% glycerol] for MtaA) by increasing concentration of imidazole. Recombinant proteins were desalted and the buffer was exchanged ([10 mM Tris, 200 mM NaCl, pH 8.0] for FAAL-PCP, FAAL-ACP, ACP and FkbH; [20 mM Bis-Tris pH 6.8, 200 mM NaCl, 10% glycerol] for MtaA) by HiPrep 26/10 Desalting column connected to HPLC system (ÄKTA pure, GE Healthcare). When SUMO-tag needs to be cut off, recombinant protein was incubated with 10% (w/w) TEV protease at 4 ^{o}C overnight. Afterwards, the SUMO-free protein was incubated with 1 mL Ni-NTA Agarose (Qiagen) at 4 ^{o}C for 1 h. The protein-resin mixture was then loaded to Ni-NTA Superflow Column (1.5 mL, Qiagen) and the gravity flow through containing tag-free protein was collected. The desalted or SUMO-free protein was purified further on Superdex 200 Increase 10/300 GL gel filtration column (for FAAL-ACP, ACP and FkbH) or HiLoad 16/600 Superdex 200 pg gel filtration column (for FAAL-PCP and MtaA) with the aid of HPLC system (ÄKTA pure). The purified proteins were concentrated with Amicon® Ultra centrifugal filter (Millipore), supplemented with 10% glycerol, snap frozen in liquid nitrogen and stored at -80 ^{o}C. The purified proteins were checked by SDS-PAGE (Fig. S9).
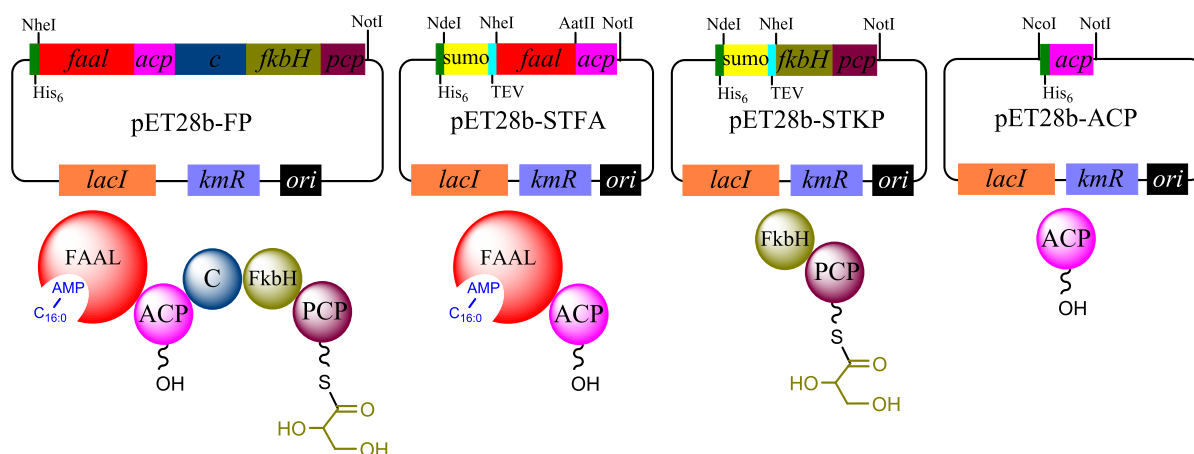
Fig. S8 The maps of protein expression vectors and the resulting domains. FAAL contained an inherent palmitoyl-AMP, whereas glycerate was loaded to PCP in *E. coli*.
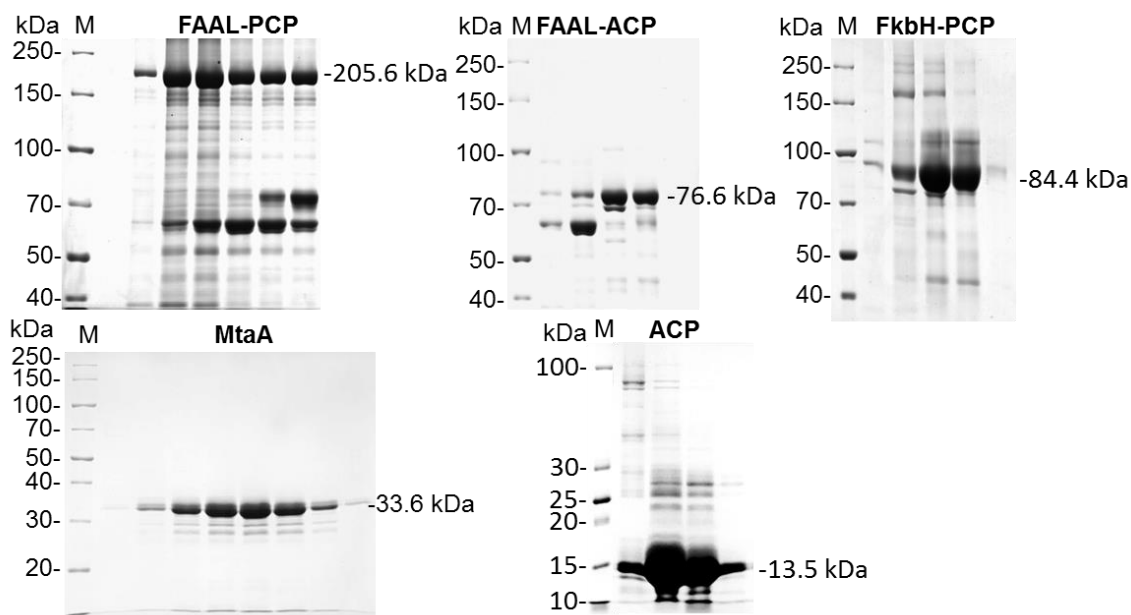

Fig. S9 SDS-PAGE analysis of the purified proteins. The fractions from gel filtration were checked by SDS-PAGE. The sizes of the purified proteins are illustrated. M, protein molecular weight (Fermentas).
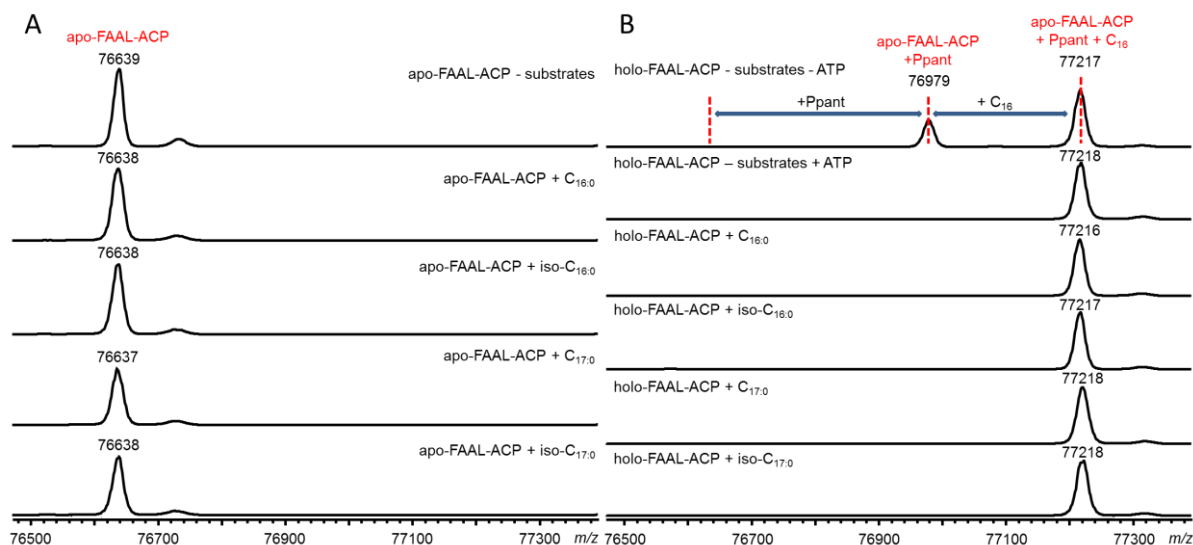
Fig. S10 Biochemical characterization of the module 1. LC-MS analysis of the fatty acid bound FAAL-ACPs. (A) fatty acids ($C_{16:0}$, *iso*-$C_{16:0}$, $C_{17:0}$ or *iso*-$C_{17:0}$) incubated with *apo*-FAAL-ACP were served as negative control; (B) fatty acids incubated with *holo*-FAAL-ACP. Only C_{16} was attached to *holo*-FAAL-ACP no matter what kind of fatty acid was added in the the reaction.

Plasmid transformation and Red/ET recombination conditions

For the routine transformation of *E. coli*, 1 mL cells ($OD_{600} \approx 0.6$) were washed twice with 1 mL ddH$_2$O and were suspended in 50 μL ddH$_2$O, 500 ng DNA was mixed well with cells in 1 mm cuvette and electroporated in the Eppendorf Eporator® at 1300 V. Transformants were recovered in 1 mL LB at 37 ^{o}C for 1 hour and were selected on LB-agar plates with appropriate antibiotics. To perform linear plus circular homologous recombination (LCHR) [2], 40 μL overnight culture of recombineering strain (*E. coli* GB05-red or GBred-gyrA462 harboring corresponding constructs) was inoculated in 1.6 mL fresh LB broth until OD_{600} reached up to 0.3, 50 μL 10% arabinose was then added to the culture to induce the expression of recombinase, and cells were allowed to grow at 37 ^{o}C for another 45 min before preparing of electro-competent cells. Electroporation of appropriate inserts into recombinase proficient cells was carried out as routine transformation.

For the transformation of myxobacteria, *M. xanthus* DK1622 (Δ*mchA*) was inoculated in CTT medium and cultivated at 30 ^{o}C for 2 days, 1.5 mL cultures was then transferred into 12 mL fresh medium and was allowed to grow at 30 ^{o}C overnight. Cells were washed twice with cold ddH$_2$O and suspended in 50 μL ddH$_2$O before being mixed with 2 μg plasmids and electroporated at 1300V. Cells were recovered in 1.6 mL CTT medium at 30 ^{o}C for 6 hours. The recovered cells were mixed with 10 mL 0.7% soft agar (with appropriate antibiotics) and plated on CTT-agar plates with kanamycin and oxytetracycline. Plates were incubated at 30 ^{o}C until single colonies appeared. For the transformation of *Burkholderia sp.* DSM 7029, 1.6 mL cells (OD$_{600}$ ≈ 0.8) was washed twice with 1.6 mL cold ddH$_2$O and suspended in 50 μL ddH$_2$O, 500 ng plasmids was mixed with cells in 1 mm cuvette and electroporation was carried out at 1300V. After recovery in 1.6 mL CYCG medium at 30 ^{o}C for 3 hours, transformants were selected on CYCG-agar plates with kanamycin. Plasmid transfer from *E. coli* GB05-red to *P. putida* KT2440 was performed by conjugation with the help of pRK2013 as previously reported [3].

Cloning and engineering of the *vio* gene cluster

The flowchart of vioprolide expression vector construction is displayed in Fig. S11. The cosmid EK21 (a) containing most part of the *vio* gene cluster was used for further engineering. The missing part of *vioZ* was amplified from the genomic DNA of *C. violaceus* Cb vi35 by using primers V-overlap-5 and vio-5, zeocin resistance gene *zeoR* was amplified with primers V-zeo and V-overlap3, an *Nhe* I restriction site was introduced in the 5'-terminus. Overlap extension PCR was used to ligate *zeoR* and *vioZ* by using primers V-overlap-5 and V-overlap3. The *zeoR-vioZ* cassette (b) was inserted into pUC-cosmid EK21 (a) by Red/ET recombination. Hereafter ampicillin resistance gene *ampR* (c) was amplified by primers V-amp-5 and V-amp-3, which replaced the *neoR*

(neomycin resistance gene) downstream of *vioD* by Red/ET recombination to generate pUC-ampR-zeoR-vioZ-cosmid EK21 (d), another *Nhe* I restriction site was introduced in front of *ampR* simultaneously. Direct cloning was used to clone the missing genes downstream of *vioD*, p15A-Tn5-kmR linear fragment (*Tn5*, *Ptn5* promoter; *IR*, inverted repeat; *kmR*, kanamycin resistance gene) with homologs of *vioD* and *orf9* was amplified by using primers vioD-15A and vioD-neo. Here, an *Avr* II restriction site was introduced in the primer vioD-15A. The p15A-Tn5-kmR linear fragment and *BsrD* I digested Cb vi35 genomic DNA were co-transformed into recombinase proficient *E. coli* GB05-red, the p15A-Tn5-kmR-vioD-orf9 (e) carring the *vioD~orf9* was then obtained by Red/ET recombination.

To stitch the *vio* gene cluster, the (d) and the (e) were digested with *Nhe* I and *Avr* II, respectively. The vioZ-cosmid EK21 fragment harboring *vioZ~vioC* was ligated with *vioD* to obtain p15A-Tn5-kmR-zeoR-vio (f). The *IR-Tps-BSD-oriT* cassette (g) (*Tps*, MycoMar transposase gene; *BSD*, blasticidin resistance gene) [4] was then amplified by using primers Vio-Tps5 and P15AUpSeq, and was integrated into the backbone of (h) by Red/ET recombination. Thereafter, *gentR-tetR-Ptet* cassette (i) was used to replace the *zeoR* in (h) to generate p15A-Tps-BSD-oriT-Tn5-kmR-gentR-tetR-Ptet-vio (j). The (j) was sequenced and aligned with Cb vi35 genome data. However, several mutation sites were discovered in *vioA*, *vioD* and *vioZ*.

Mutation sites in the *vio* gene cluster were repaired by *ccdB* counter selection as previously reported [5]. In order to repair the mutation site in *vioD*, *vioDampR-ccdB* cassette (k) was amplified from p15A-ccdB-amp and transformed into recombineering competent *E. coli* GBred-gyrA462 harboring (j), recombinants (l) were selected on LB-agar plates with kanamycin and ampicillin. The *vioD-3term* fragment containing 201 bp 3'-terminus of *vioC* was amplified from Cb vi35 genome by using primers vioCDlig15 and vioCDlig13,

and the *vioD5term* fragment contains 204 bp 5'-termini of *vioD* was amplified from the genome by using the primers vioCDlig25 and vioCDlig23. The *vioC-3term* and the *vioD5term* were ligated by overlap extension PCR to obtain *vioCDlig* (m), 500 ng (m) was transformed into recombineering proficient *E. coli* GBred-gyrA462 harboring (l) to replace the *ampR-ccdB* cassette, cells were recovered in 1.6 mL LB medium for 3 hours, the plasmids were then extracted by alkaline lysis and retransformed into *E. coli* GB05-red, *E. coli* GB05-red harboring p15A-Tps-BSD-oriT-Tn5-kmR-gentR-tetR-Ptet-violig (n) were selected on LB-agar plates supplemented with kanamycin. Recombinants were checked by restriction enzymes and sequencing.

To repair the mutation site in *vioA*, *vioAcmR-ccdB* cassette (o) was amplified from p15A-ccdB-Cm by using primers XmaJIhomolog-cm5 and Homolog-cmccdB3, and was transformed into recombineering proficient *E. coli* GBred-gyrA462 in which (n) was resident. Recombinants (p) were selected on LB-agar plates with kanamycin and chloramphenicol. The 466 bp *vioAseq2* (q) with corrected mutation site of *vioA* was amplified from the genome of *C. violaceus* Cb vi35 by using primers vioAseq5 and vioAseq32. After transforming 500 ng (q) into recombinase induced *E. coli* GBred-gyrA462 harboring (p), *cmR-ccdB* cassette was removed and plasmids were transformed into *E. coli* GB05-red. *E. coli* GB05-red carring the vioprolide expression construct p15A-Ptet-vioArecov-vio (r) was selected on LB-agar plate with kanamycin. Recombinants were analyzed by *Pvu* II and verified further by Sanger-sequencing (Fig. S12).

To repair the mutation site in *vioZ*, the primers Ptetviozcmccdb5 and Ptetviozcmccdb3 were used to amplify *vioZcmR-ccdB* (s) from p15A-ccdB-Cm. The PCR product was transformed into recombinase proficient *E. coli* GBred-gyrA462 harboring (r) and was inserted between *Ptet* promoter and *vioZ* to generate (t) by Red/ET recombination. Correct 3'-terminus of *vioZ* (u) was amplified from the genome of *C. violaceus* Cb vi35 by using the primers

Ptetviozrecov5 and vioZrecov3, which then replaced the *cmR-ccdB* cassette to obtain (v). In this way, the *Ptet* promoter was retained to drive the whole gene cluster. In order to change the *Ptet* promoter to *Ptn5* promoter, *vioZampR-ccdB* cassette (w) was amplified from p15A-ccdB-amp by using the primers vioZampccdB5 and vioZampcounter3. The PCR product was transformed into recombinase proficient *E. coli* GBred-gyrA462 harboring (r) to replace the *gentR-tetR-Ptet* cassette. Correct 5'-terminus of *vioZ* (y) was amplified from the genome of *C. violaceus* Cb vi35 by using primers vioZrecov5 and vioZrecov3, and substituted the *ampR-ccdB* cassette *via* Red/ET recombination to generate (z). In this case, *vioZ* was co-transcripted with *kmR* and was controlled by *Ptn5* promoter. The vioprolide final expression vectors p15A-Ptet-vio (v) and p15A-Tn5-vio (z) were verified by restriction digestions and Sanger-sequencing (Fig. S13).
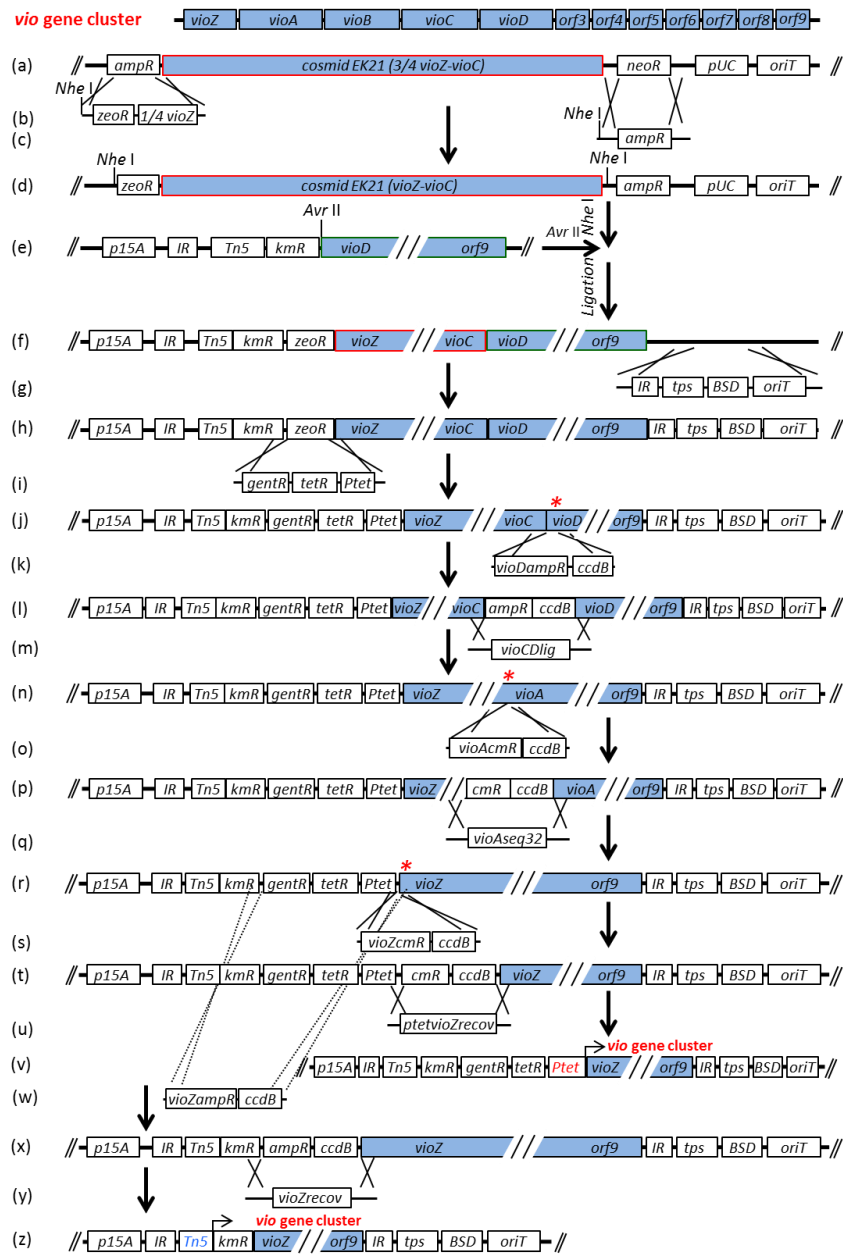
Fig. S11 Engineering of *vio* gene cluster for heterologous expression. The *vio* gene cluster was cloned from cosmid library and was constructed to expression vector by conventional cloning (a-f); elements for transcription and transposition were added to expression vector (g-i); several mutation sites were repaired by *ccdB* counter selections (j-z). *, mutation sites discovered in the *vio* gene cluster.
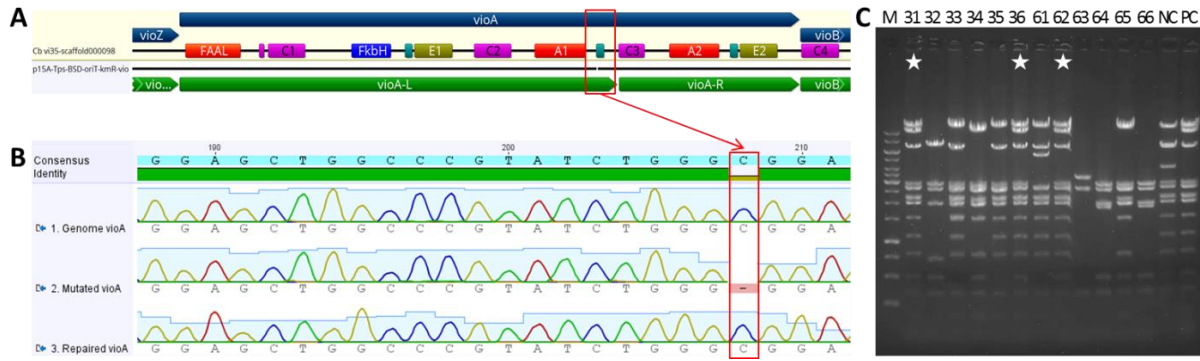
Fig. S12 Repair of the mutation site in *vioA*. The mutation site in *vioA* is indicated with a red box (A). The repaired site was verified by Sanger-sequencing (B) and *Pvu* II digestion (C). The correct clones are indicated with white stars. NC, negative control (p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-vioAcmccdB-violig); PC, positive control (p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-violig); M, 1kb DNA ladder.
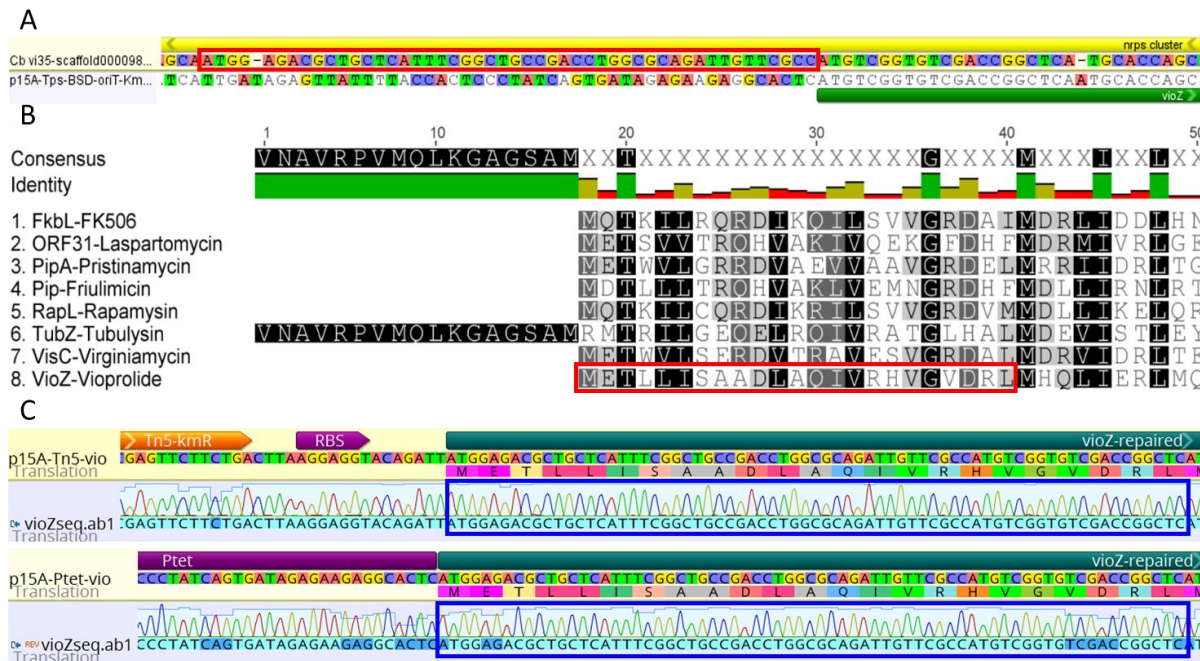

Fig. S13 Identification and repair of the missing part in *vioZ*. The missing part of *vioZ* was identified by comparing the DNA sequence with the genome sequence of *C. violaceus* Cb vi35 (A) or by alignment of the protein sequence with other cyclodeaminase-like proteins (B); the repaired *vioZ* was verified by Sanger-sequencing (C). The missing part of *vioZ* is indicated with red box; the repaird part is indicated with blue box.
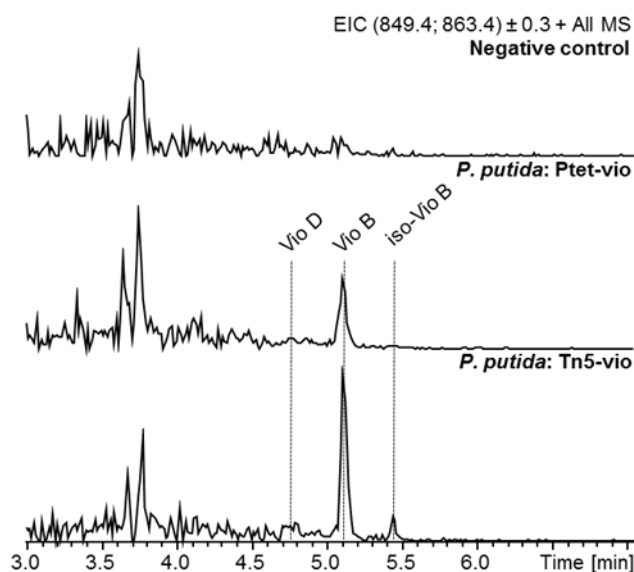
Fig. S14 LC-HRMS analysis of vioprolide production in *P. putida* KT2440. Extracted ion chromatogram (EIC) *m/z* for 849.4 [M+H]$^+$ (vioprolide D) and 863.4 [M+H]$^+$ (vioprolide B and isomer) is shown. Intensity is set to the same range.

Generation of Δ*orf3*-Δ*orf9* mutants

The genes *orf3-orf9* in the downstream of *vioD* were deleted by Red/ET recombination. The flow chart is shown in Fig. S15. Nine pairs of primers (orf3Del-cm5 / orf3Del-cm3, orf4Del-cm5 / orf4Del-cm3, orf5Del-cm5 / orf5Del-cm3, orf6Del-cm5 / orf6Del-cm3, orf5Del-cm5 / orf6Del-cm3, orf7Del-cm5 / orf7Del-cm3, orf8Del-cm5 / orf8Del-cm3, orf9Del-cm5 / orf9Del-cm3, and orf3Del-cm5 / orf9Del-cm3,) containing the respective homologs of *orf3* to *orf9* were used to amplify *cmR* from p15A-ccdB-Cm. *XmaJ* I restriction sites were designed in the primers. The resulting inserts *orf3Delcm*, *orf4Delcm*, *orf5Delcm*, *orf6Delcm*, *orf5-6Delcm*, *orf7Delcm*, *orf8Delcm*, *orf9Delcm* and *orf3-9Delcm* were transformed into recombinase proficient *E. coli* GBred::p15A-Ptet-vioArecov-vio, and the *orf3*, *orf4*, *orf5*, *orf6*, *orf5&6*, *orf7*, *orf8*, *orf9* or *orf3-9* on p15A-Ptet-vioArecov-vio was replaced by *cmR* via Red/ET recombination, respectively. The resulting p15A-Ptet-vioArecov-vio-

orfNDelcmR (N=3, 4, 5, 6, 5&6, 7, 8, 9 or 3-9) was digested with *XmaJ* I to remove the *cmR*, and the linear construct was self-ligated by T4 DNA ligase. The mutants, designated p15A-Ptet-vioArecov-vio-orfNDel (N=3, 4, 5, 6, 5&6, 7, 8, 9 or 3-9), were checked by *Bsp119* I *and XmaJ* I (Fig. S16), and were transformed into *M. xanthus* DK1622 (Δ*mchA*) and *Burkholderia* DSM 7029. The transformants were checked by colony PCR (Fig. S17). The *M. xanthus* transformants were fermented in CTT medium, whereas *Burkholderia* transformants were fermented in CYCG medium. Compounds were extracted with XAD16 and methanol. The production of vioprolides was measured with LC-MS (Fig. S18).
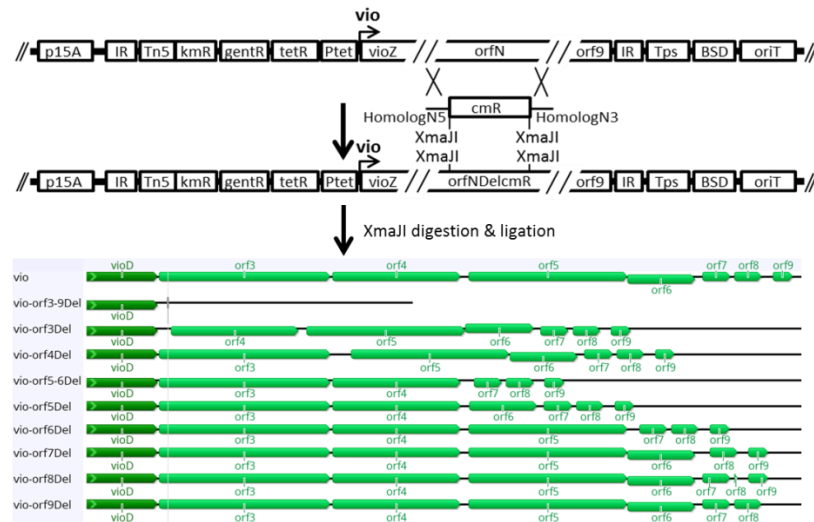

Fig. S15. Flow chart for the construction of Δ*orf3*-Δ*orf9* mutants. The *orfN* (N=3, 4, 5, 6, 5&6, 7, 8, 9 or 3-9) was initially replaced by *cmR* via Red/ET recombination, the resulting constructs were then digested with *XmaJ* I to remove *cmR* and were self-ligated with T4 DNA ligase.
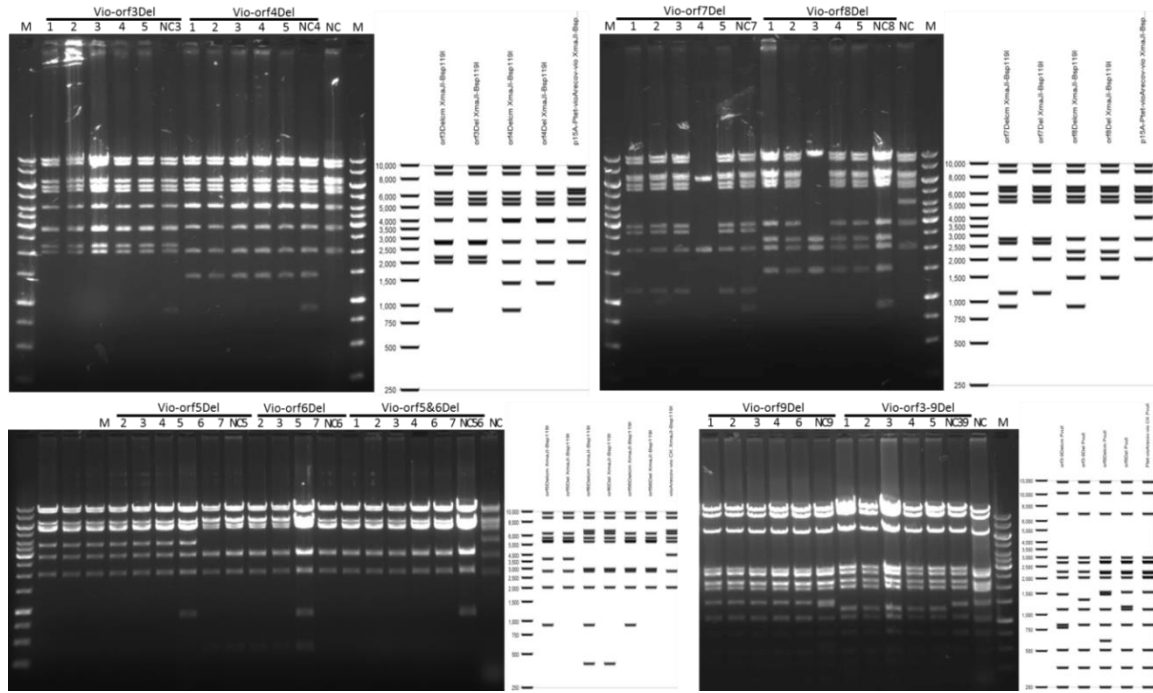
Fig. S16 Restriction analyses of Δ*orf3*-Δ*orf9* mutants. The mutants p15A-Ptet-vioArecov-vio-orfNDel (N=3, 4, 5, 6, 5&6, 7, 8, 9 or 3-9) were checked with *Bsp119* I and *XmaJ* I; p15A-Ptet-vioArecov-vio-orfNDelcm (NC3-NC9) and p15A-Ptet-vioArecov-vio (NC) were set as negative control; M, 1kb DNA ladder.
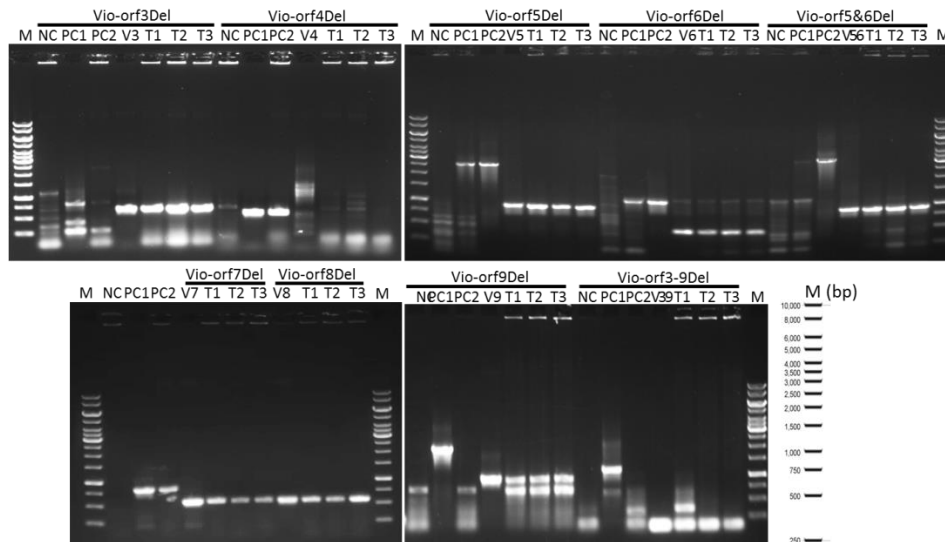

Fig. S17 Colony PCR analyses of Δ*orf3*-Δ*orf9* transformants. NC, *M. xanthus* DK1622; PC1, p15A-Ptet-vioArecov-vio; PC2, *M. xanthus*::Ptet-vioArecov-vio; V3-V9, plasmid control, p15A-Ptet-vioArecov-vio-orfNDel (N=3, 4, 5, 6, 5&6, 7, 8, 9 and 3-9, respectively); T1-T3, *M. xanthus* transformants; M, 1kb DNA ladder.
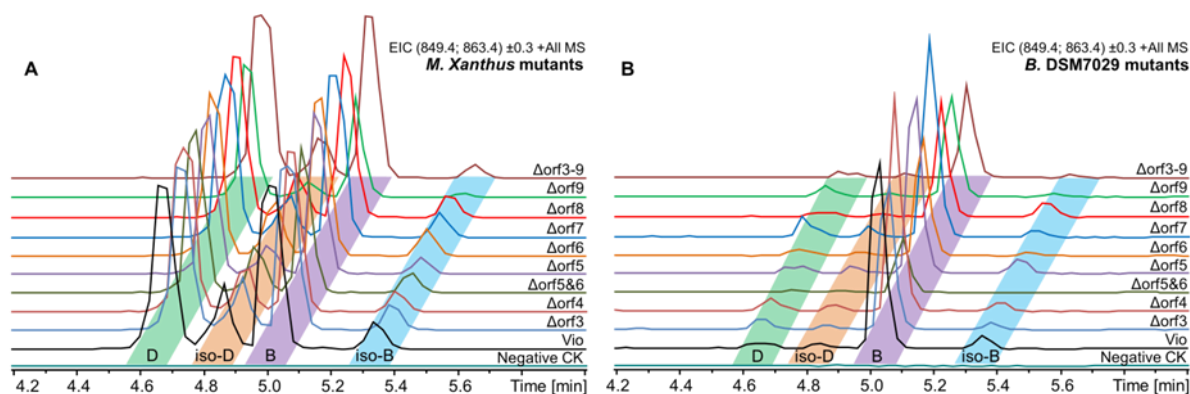
Fig. S18 HPLC-MS analyses of vioprolides in Δ*orf3*-Δ*orf9* mutants. (A) production of vioprolides in *M. xanthus* DK1622 (*ΔmchA*) transformants; (B) production of vioprolides in *Burkholderia sp.* DSM7029 transformants. Extracted ion chromatograms (EIC) *m/z* for 849.4 $[M+H]^+$ (vioprolide Dand isomer) and 863.4 $[M+H]^+$ (vioprolide B and isomer) are shown. The produced vioprolides are indicated with shadowed boxes. Intensity is adjusted to the same range.

Site-directed mutagenesis on A_1 and NMT domain

Site-directed mutagenesis was performed by *ccdB* counter selection. To mutate the position 331 (Glu) of the A_1 domain into Asp and the conservative Gly of the NMT domain into Arg, the *cmR-ccdB* cassette with homologs of A_1 domain and NMT domain were amplified from p15A-ccdB-Cm by using primer pairs VioA1mutcmccdB5 / VioA1mutcmccdB3 and NMTmutcmccdB5 / NMTmutcmccdB3, respectively. The PCR products *VioA1mutcmccdB* or *NMTmutcmccdB* was transformed into recombinase proficient *E. coli* GBred-gyrA462::p15A-Ptet-vioArecov-vio, and the target region of A_1 or NMT domain were replaced by *cmR-ccdB*, respectively. The recombinants were selected on LB-agar plates with kanamycin and chloramphenicol. The site-mutated fragments *VioA1E-D* and *NMT-G-R* were amplified from p15A-Ptet-vioArecov-vio by using primer pairs VioA1E-D5 / VioA1E-D3 and NMT-G-R-5 / NMT-G-R-3, and were transformed into recombinase proficient *E. coli* GBred-gyrA462::p15A-Ptet-vioArecov-vioA1mutcmccdB and *E. coli* GBred-

gyrA462::p15A-Ptet-vioArecov-NMTmutcmccdB, respectively. The resulting recombinants were designated p15A-Ptet-vioArecov-A1E331D and p15A-Ptet-vioArecov-NMTmut, respectively. The A_1 domain and NMT domain double mutation was generated by introducing A_1E331D mutation in p15A-Ptet-vioArecov-NMTmut in the same way. All clones were verified by restriction analysis and Sanger-sequencing (Fig. S19).
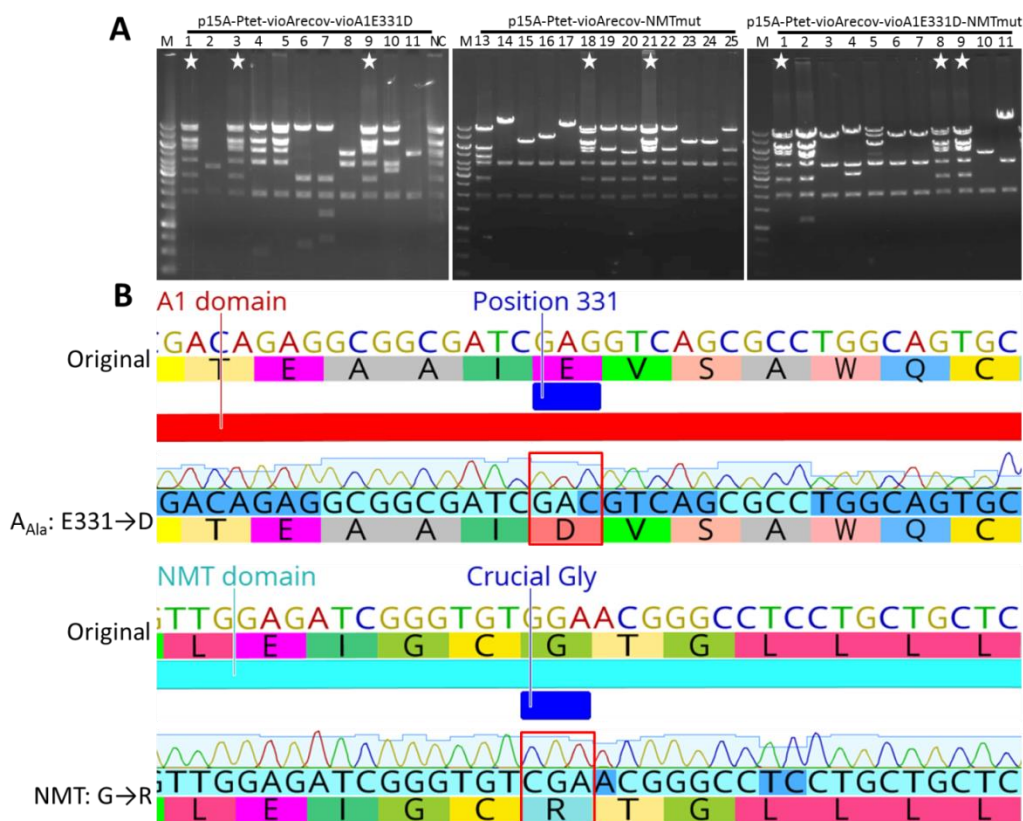
Fig. S19 Anaysis of the site-directed mutagenesis on A_{Ala} domain and NMT domain. Mutants were checked by *Bsp119* I and *XmaJ* I (A) and Sanger-sequencing (B). The correct clones are indicated with white stars; the position 331 of the A_1 domain and the crucial Gly in the NMT domain are illustrated; the mutated positions are labeled with red boxes.

Purification of vioprolide derivatives and structure elucidation

M. xanthus::Ptet-vioArecov-vio or *M. xanthus*::Ptet-vioArecov-vioA1mut was cultivated in 5 × 2 L CTT medium at 30 °C for 4.5 days, 33.4 mg/L pipecolic acid was added every 24 hours, 1% Ambliter XAD-16 resin was added after 4

days and was incubated with cultures overnight. Compounds were extracted with methanol and the solvent was evaporated with a vacuum system. The dried substance was extracted with ethyl acetate : ddH2O (2 : 1) and was concentrated by evaporation. The derivatives were separated with Sephadex LH-20 column (GE Healthcare) using methanol as mobile phase. Fractions were collected with a sample collector. Fractions containing vioprolide derivatives were purified further by HPLC.

Purification of vioprolides was carried out on a Dionex (Germering, Germany) Ultimate 3000 low pressure gradient system, equipped with SR3000 solvent rack, LPG-3400SD pump module, WPS-3000TSL autosampler, Knauer (Berlin, Germany) Jetstream column oven and a DAD-3000 photodiode array detector with a cell volume of 10 μL. Cleanup was carried out using 100 μL injections of sephadex fractions reconstituted in 250-400 μL MeOH on a Phenomenex Luna C$_{18}$(2) HPLC column (250 $\times$ 10 mm, 4 μm). The gradient conditions were as follows: 0-52 min, 5-80% B; 52-52.5 min, 80-95% B; 52.5-56 min, 95% B; 56-56.5 min, 95-5% B; 56.5-66.5 min, 5% B. Mobile phases were (A) water and (B) methanol. The separation was carried out with a column temperature of 40 °C at a flow rate of 5 mL/min. Fractions were collected using an AFC-3000 fraction collector based on retention time. Impure fractions were further purified using 100 μL injections of LC fractions reconstituted in 350-1000 μL MeOH on an Agilent Zorbax XDB C8 HPLC column (9.4 $\times$ 250 mm, 5 μm). The following gradient conditions were used: 0-40 min, 30-85% B; 40-40.1 min, 85-95% B; 40.1-42.1 min, 95% B; 42.1-42.3 min, 95-30% B; 42.3-52 min, 30% B. Mobile phases were (A) water and (B) methanol. The separation was carried out with a column temperature of 40 °C at a flow rate of 5 mL/min. Fractions were collected using an AFC-3000 fraction collector based on retention time.

2. Manuscript

Purification of hydroxyl-vioprolides was carried out on a Dionex (Germering, Germany) Ultimate 3000 low pressure gradient system, equipped with SR3000 solvent rack, LPG-3400SD pump module, WPS-3000TSL autosampler, Knauer (Berlin, Germany) Jetstream column oven and a DAD-3000 photodiode array detector with a cell volume of 10 μL. LC separations were performed using 100 μL injections on a Agilent Zorbax XDB C8 HPLC column (9.4 × 250 mm, 5 μm) at 40°C and a flow rate of 5 mL/min with the following gradient conditions: 0-40 min, 10-65 %B; 40-40.1 min, 65-95% B; 40.1-44.1 min, 95% B; 44.1-44.3 min, 95-10% B; 44.3-54 min, 10% B. Mobile phases were (A) water and (B) methanol.

For the purification of previoprolides, initial cleanup was carried out by using 500 μL injections of methanolic cell extract on a Phenomomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 μm). A linear gradient with the following conditions was used: 0-20 min, 5-95%B; 20-22 min, 95%B; 22-22.5 min, 95-5% B; 22.5-28.5 min, 5% B. Mobile phases were (A) water and (B) acetonitrile. The separation was carried out with a column temperature of 40 °C at a flow rate of 7 mL/min. And the final purification step was carried out by using 500 μL injections of crude fractions reconstituted in 6 mL methanol on an Agilent Zorbax XDB C$_8$ HPLC column (9.4 × 250 mm, 5 μm). A linear gradient with the following conditions was used: 0-7 min, 5-75% B; 7-35 min, 75-95% B; 35-38 min, 95% B; 38-38.5 min, 95-5% B; 38.5-47 min, 5% B. Mobile phases were (A) water and (B) acetonitrile. The separation was carried out with a column temperature of 40 °C at a flow rate of 7 mL/min. Fractions were collected using an AFC-3000 fraction collector based on retention time.

Structures of the purified vioprolides and vioprolide D1 were assigned with NMR (Table S3). The NMR spectra of the vioprolide derivatives were recorded in methanol-*d4* on a Bruker Avance III 500 NMR spectrometer (Bruker Karlsruhe, Germany) with a 5 mm TCI cryoprobe (^{1}H at 500 MHz, ^{13}C at 125

MHz) or a Bruker Ascend 700 spectrometer with a 5 mm TXI cryoprobe (^{1}H at 700 MHz, ^{13}C at 175 MHz). Chemical shifts were calibrated internally to the residual signal of methanol-*d4* in which the sample was dissolved (δH 3.31, δC 49.15).

Table S3 NMR spectroscopic data of vioprolide D, vioprolide D isomer and vioprolide D_1

		vioprolide D		vioprolide D (isomer)		vioprolide D_1	
		δ_C	δ_H	δ_C	δ_H	δ_C	δ_H
N-methylvaline	1	171,6		170,9		171,8	
	2	67,3	4,08	66,7	4,53	67,8	4,13
	3	28,7	2,34	29,3	2,32	28,7	2,37
	4	20,7	1,01	19,9	1,07	20,7	1,04
	5	19,7	0,90	19,9	1,01	19,7	0,91
	N-Me	**36,3**	**3,32**	**29,9**	**2,91**	36,6	3,24
	O-Me						
threonine	1	172,7		173,4		172,7	
	2	55,4	4,79	55,8	4,85	56,1	4,77
	3	68,4	4,15	68,6	4,07	68,4	4,13
	4	19,9	1,14	20,3	1,19	20,2	1,17
proline/ homoproline	1	173,6		174,5		173,8	
	2	61,9	4,58	61,9	4,56	61,9	4,60
	3a	31,0	2,29	31,1	2,29	30,7	2,20
	3b		2,15		2,09		2,26
	4	25,2	1,96	25,4	1,94	25,2	1,96
	5a	49,2	3,57	49,1	3,50	49,3	3,52
	5b		3,51		3,45		3,63
	6a						
	6b						
dehydrobutyrine	1	166,6		166,6		167,0	
	2	131,9		131,8		131,8	
	3	118,2	5,82	118,4	5,98	118,6	5,82
	4	12,9	1,72	13,1	1,71	13,0	1,73
cysteine	1	170,3		171,2		170,9	
	2	79,0	5,15	78,2	5,16	78,6	5,12
	3a	36,0	3,68	38,9	3,72	36,6	3,50
	3b		3,48		3,67		3,74
4-methylazetidine-carboxylic acid	1	182,2		178,9		182,8	
	2	61,1	4,90	61,0	4,75	61,2	4,93
	3a	31,3	2,33	31,2	2,28	31,2	2,33

|							
	3b		2,04		2,10		2,06
	4a	25,6	2,24	24,9	2,07	25,7	2,23
	4b		2,08				2,07
	5a	48,3	3,98	48,3	4,06	48,2	3,96
	5b		3,68		3,91		3,65
leucine	1	173,2		173,9		173,2	
	2	50,6	4,69	51,0	4,70	50,7	4,75
	3a	40,8	1,60	40,9	1,64	40,8	1,58
	3b		1,50		1,53		1,51
	4	25,7	1,69	25,9	1,71	25,7	1,69
	5	23,5	0,97	23,5	0,99	23,5	0,97
	6	21,8	0,95	21,8	0,97	22,1	0,96
alanine	1	173,9		173,9		171,7	
	2	50,1	4,40	49,7	4,40	56,1	4,45
	3	18,8	1,38	19,3	1,33	**63,1**	**3,82**
						(Serine β-OH)	
glyceric acid	1	172,9		172,4		173,6	
	2	70,8	4,41	71,6	4,39	71,1	4,50
	3a	66,1	4,41	66,9	4,56	66,7	4,47
	3b		4,09		4,18		4,11
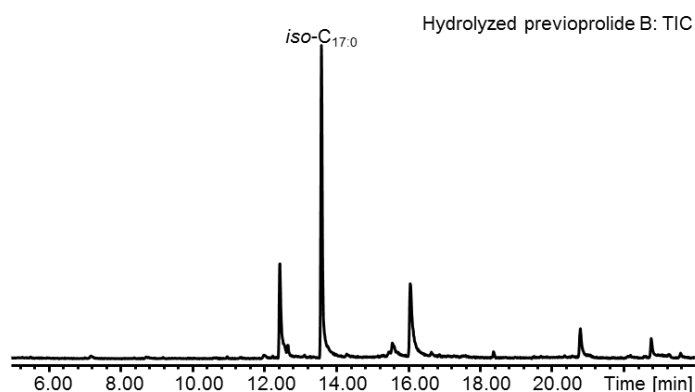
Fig. S20 GC-MS analysis of the fatty acid component in previoprolide B. The total ion current (TIC) chromatogram is shown.

Quantification of vioprolides

Aliquots of 100 μL were withdrawn from 25 mL bacterial cultures over a time course of 14 days. Each strain was cultivated in triplicates. Withdrawn sample portions were centrifuged for 10 min at 21000 g and 4 °C by using a Himac CT15E benchtop centrifuge (Hitachi Koki Co., Ltd., Japan), followed by

100

withdrawal of 80 μL of supernatant. These 80 μL samples were dried using an Eppendorf Concentrator plus (Hamburg, Germany) and the dried down samples were stored at -20 °C until measurement. Prior to measurement, all samples were reconstituted in 80 μL of MeOH in conical LC vials. All measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 × 2.1 mm, 1.7 μm dp column (Waters, Germany). Separation of 5 μL sample was achieved by a linear gradient from (A) H_2O + 0.1% fomic acid to (B) ACN + 0.1% formic acid at a flow rate of 600 μL/min and a column temperature of 45 °C. Gradient conditions were set as follows: 0-0.5 min, 5% B; 0.5-9.5 min, 5-45% B; 9.5-10.5 min, 45-95% B; 10.5-11.5 min, 95%; 11.5-11.8 min, 95-5% B; 11.8-13.5 min, 5% B. The eluent stream was diverted to waste during the first 7 min in order to minimize source contamination. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 μL/min before entering the solariX XR (7T) FT-ICR mass spectrometer (Bruker Daltonics, Germany) using the Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 0 V during the first 7 min and 4500 V after 7 min, the dry-gas flow was 4.0 L/min and the nebulizer was set to 1.1 bar. After the generated ions passed the quadrupole with a low cutoff at 100 *m/z* they were trapped in the collision cell for 200 ms and finally transferred within 0.9 ms through the hexapole into the ICR cell. Captured ions were excited by applying a frequency sweep from 100 to 1600 *m/z* and detected in broadband mode using a FID data size of 256k. Transient length was intentionally kept low to ensure a sufficient number (>14) of points per peak for quantification purposes.

The calibration curves for external calibration were obtained using solutions of vioprolide B and D with concentrations in the range 100-0.01 ng/mL. Curve fitting was done using the Origin 2016 software package (OriginLab, Northampton, UK). Regression analysis was done using both linear and

quadratic regression models; a subsequent F-Test for model comparison was carried out to identify the most suitable model. A quadratic model was found to be appropriate for both analytes ($\alpha=0.05$). Error bars are prediction intervals of the mean calculated concentrations obtained from three biological replicates at the indicated incubation times. Calculation of prediction intervals and limits of quantification was achieved using a custom Microsoft Excel 2010 sheet. The limit of quantification was iteratively calculated by employing the GRG nonlinear solving method with a convergence criterion of $2\cdot10^{-5}$. The production titers of vioprolide B and D from Cb vi35 and heterologous hosts are shown in Fig. S21.
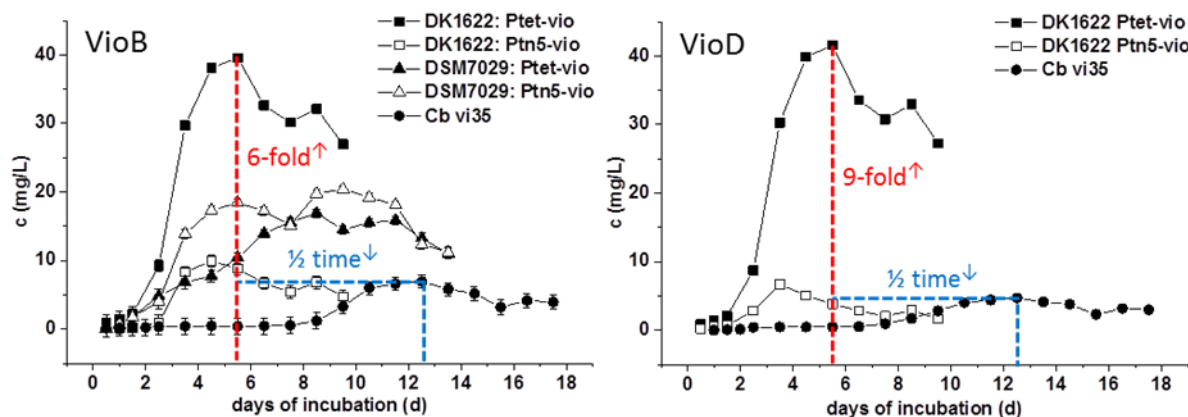

Fig. S21 Quantification of vioprolide B and D from Cb vi35 and heterologous hosts.

Strains and primers

Table S4 Strains and primers used in this study

Strains	Characterization
E. coli GB2005	*E. coli* strain for cloning, derivate strain of *E. coli* DHB101
E. coli GB05-red	*E. coli* GB2005 with *redα*, *redβ*, *redγ* and *recA* integrated into chromosome under control of arabinose inductive P_{BAD} promoter [6]
E. coli GBred-gyrA462	GB05-red, GyrA mutation of Arg462Cys [5]
E. coli HS996::pUC-cosmid EK21	*E. coli* HS996 carring the cosmid vector pUC-cosmid EK21; most part of *vio* gene cluster was located in cosmid EK21; Amp^R
E. coli GBred::pUC-zeoR-vioZ-cosmid EK21	*E. coli* GB05-red carring pUC-zeoR-vioZ-cosmid EK21; zeoR-vioZ cassette was inserted on pUC-cosmid EK21, the *vioZ* gene is integral; Neo^R, Zeo^R
E. coli GBred::pUC-ampR-zeoR-vioZ-cosmid EK21	The *neoR* gene of pUC-zeoR-vioZ-cosmid EK21 was replaced by *ampR*; Amp^R, Zeo^R
E. coli GBred::p15A-Tn5-kmR-	*E. coli* GB05-red carring p15A-Tn5-kmR-vioD~orf9; *vioD~orf9* were cloned on

vioD~orf9	p15A vector; Km^R
E. coli GBred::p15A-Tn5-kmR-zeoR-vio	*E. coli* GB05-red carring p15A-Tn5-kmR-zeoR-vio; the *vio* gene cluster was stitched and constructed on p15A vector; Km^R, Zeo^R
E. coli GBred::p15A- Tps-BSD-oriT- kmR-zeoR-vio	*E. coli* GB05-red carring p15A-Tps-BSD-oriT- kmR-zeoR-vio; the Tps-BSD-oriT cassette was added on p15A-Tn5-kmR-zeoR-vio; Km^R, Zeo^R, BSD^R
E. coli GBred::p15A-Tps-BSD-oriT-KmR-gentR-tetR-Ptet-vio	*E. coli* GB05-red carring p15A-Tps-BSD-oriT-KmR-gentR-tetR-Ptet-vio on which *gent^R-tet^R-Ptet* was inserted before *vioZ* where *zeoR* was replaced; *vioZ*, *vioA* and *vioD* were mutated. Km^R, Genta^R, BSD^R
E. coli GBred-gyrA462::p15A-Tps-BSD-oriT-Km-gentR-Ptet-vio-ampccdB-vioD	*E. coli* GBred-gyrA462 carring p15A-Tps-BSD-oriT-KmR-gentR-tetR-Ptet-vio-ampccdB-vioD, the mutation site in *vioD* was substituted by *ampR-ccdB*; Km^R, Genta^R, Amp^R, BSD^R
E. coli GBred::p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-violig	*E. coli* GB05-red carring p15A-Tps-BSD-oriT-KmR-gentR-tetR-Ptet-violig, the *vioD* was repaired; Km^R, Genta^R, BSD^R
E. coli GBred-gyrA462::p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-violig	*E. coli* GBred-gyrA462 carring p15A-Tps-BSD-oriT-KmR-gentR-tetR-Ptet-violig; Km^R, Genta^R, BSD^R
E. coli GBred-gyrA462::p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-vioAcmccdB-violig	*E. coli* GBred-gyrA462 carring p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-vioAcmccdB-violig, the mutation site of *vioA* gene was replaced by *cmR-ccdB* cassette; Km^R, Genta^R, Cm^R, BSD^R
E. coli GBred::p15A-Ptet-vioArecov-vio	*E. coli* GB05-red carring the final vioprolides expression vector p15A-Tps-BSD-oriT-Tn5-KmR-gentR-tetR-Ptet-vioArecov-vio, the mutation site in *vioA* was repaired; Km^R, Genta^R, BSD^R
E. coli GBred- gyrA462::p15A-Tps-BSD-oriT-Tn5-Km-vioZampccdB-vioArecov-vio	*E. coli* GBred- gyrA462 carring p15A-Tps-BSD-oriT-Tn5-KmR-vioZampccdB-vioArecov-vio, the *gentR-tetR-Ptet* cassette and the mutation site in *vioZ* were replaced by *ampR-ccdB* cassette; Km^R, BSD^R, Amp^R
E. coli GBred:: p15A-Ptn5-vio	*E. coli* GB05-red carring the final vioprolides expression vector p15A-Tps-BSD-oriT-Tn5-KmR-vio, the mutation sites in both *vioZ* and *vioA* were repaired; *vio* gene cluster was controlled by *Ptn5* promoter; Km^R, BSD^R
E. coli GBred- gyrA462::p15A-Tps-BSD-oriT-Tn5-Km-gentR-Ptet-vioZcmccdB-vioArecov-vio	*E. coli* GBred-gyrA462 carring p15A-Tps-BSD-oriT-Tn5-KmR-gentR-tetR-Ptet-vioZcmccdB-vioArecov-vio, the mutation site in *vioZ* was replaced by *cmR-ccdB* cassette; Km^R, Genta^R, BSD^R, Cm^R, BSD^R
E. coli GBred:: p15A-Ptet-vio	*E. coli* GB05-red carring vioprolide final expression vector p15A-Tps-BSD-oriT-Tn5-KmR-gentR-tetR-Ptet-vio, the mutation sites in *vioZ*, *vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptet* promoter; Km^R, Genta^R, BSD^R
E. coli GBred::p15A-Ptet-vioArecov-vio-orfNDelcmR	*E. coli* GB05-red carring p15A-Ptet-vioArecov-vio-orfNDelcmR (N= 3, 4, 5, 6, 5&6, 7, 8, 9 or 3~9); the *orf3, orf4, orf5, orf6, orf5&6, orf7, orf8, orf9 or orf3~9* in the *vio* gene cluster was replaced with *cmR* gene; Km^R, Genta^R, Cm^R, BSD^R
E. coli GBred::p15A-Ptet-vioArecov-vio-orfNDel	*E. coli* GB05-red carring p15A-Ptet-vioArecov-vio-orfNDel (N= 3, 4, 5, 6, 5&6, 7, 8, 9 or 3~9); the *orf3, orf4, orf5, orf6, orf5&6, orf7, orf8, orf9 or orf3~9* in the *vio* gene cluster was deleted; Km^R, Genta^R, BSD^R
E. coli GBred-gyrA462::p15A-Ptet-vioArecov-vioA1mutcmccdB	*E. coli* GBred-gyrA462 carring p15A-Ptet-vioArecov-vioA1mutcmccdB; the Glu (position 331) encoding codon GAG in the A$_1$ domain was replaced by *cmR-ccdB* cassette; Km^R, Genta^R, Cm^R, BSD^R
E. coli GBred::p15A-Ptet-vioArecov-vioA1E331D	*E. coli* GB05-red carring p15A-Ptet-vioArecov-vioA1E331D; the position 331 in the A$_1$ domain was mutated to Asp; Km^R, Genta^R, BSD^R
E. coli GBred-gyrA462::p15A-Ptet-vioArecov-NMTmutcmccdB	*E. coli* GBred-gyrA462 carring p15A-Ptet-vioArecov-NMTmutcmccdB; the conservative Gly encoding codon GGA was replaced by *cmR-ccdB* cassette; Km^R, Genta^R, Cm^R, BSD^R
E. coli GBred::p15A-Ptet-vioArecov-NMTmut	*E. coli* GB05-red carring p15A-Ptet-vioArecov-NMTmut; the NMT domain in *vio* gene cluster was inactivated; Km^R, Genta^R, BSD^R
E. coli GBred-gyrA462::p15A-Ptet-vioArecov-vioA1mutcmccdB-NMTmut	*E. coli* GBred-gyrA462 carring p15A-Ptet-vioArecov-vioA1mutcmccdB-NMTmut; the Glu (position 331) encoding codon GAG in the A$_1$ domain was replaced by *cmR-ccdB* cassette; the NMT domain in *vio* gene cluster was inactivated; Km^R, Genta^R, Cm^R, BSD^R
E. coli GBred::p15A-Ptet-vioArecov-vioA1E331D-NMTmut	*E. coli* GB05-red carring p15A-Ptet-vioArecov-NMTmut; the position 331 in the A$_1$ domain was mutated to Asp and NMT domain was inactivated; Km^R, Genta^R, BSD^R
E. coli BL21 (DE3)::pCold-MtaA + pGro7	MtaA expressing strain; kindly provided by Dr. Hilda Sucipto; Amp^R, Cm^R
E. coli Rosetta (DE3)::pET28b-STFP	*E. coli* Rosetta (DE3) carring pET28b-STFP; FAAL-ACP-C-FkbH-PCP expression strain; *faal-acp-c-fkbH-pcp* with a sumo-tag and a TEV-protease cleavage site was constructed on expression vector pET28b; Km^R
E. coli Rosetta (DE3)::pET28b-FP	*E. coli* Rosetta (DE3) carring pET28b-FP; FAAL-ACP-C-FkbH-PCP expression strain; *faal-acp-c-fkbH-pcp* was constructed on expression vector pET28b; Km^R
E. coli Rosetta (DE3)::pET28b-	*E. coli* Rosetta (DE3) carring pET28b-STFA; FAAL-ACP expression strain; *faal-acp*

STFA	with a sumo-tag and a TEV-protease cleavage site was constructed on expression vector pET28b; Km^R
E. coli BL21 (DE3)::pET28b-ACP	*E. coli* BL21 (DE3) carring pET28b-ACP; ACP expression strain; ACP encoding sequence was constructed on expression vector pET28b; Km^R
E. coli BL21 (DE3)::pET28b-STKP	*E. coli* BL21 (DE3) carring pET28b-STKP; FkbH-PCP expression strain; *fkbH-pcp* with a sumo-tag and a TEV-protease cleavage site was constructed on expression vector pET28b; Km^R
E. coli GBred::p15A-Ptet-vio-FACdel	*E. coli* GBred carring p15A-Ptet-vio-FACdel, the *faal-acp-c* tridomain of the *vio* gene cluster was deleted; Km^R
E. coli GBred-gyrA462:: p15A-Ptet-vio-C4cmccdB	*E. coli* GBred-gyrA462 carring p15A-Ptet-vio-C4cmccdB; the catalytic motif of C$_4$ domain was replaced by *cmR-ccdB* cassette; Cm^R, Km^R
E. coli GBred:: p15A-Ptet-vio-C4H1mut	*E. coli* GB05-red carring p15A-Ptet-vio-C4H1mut; the encoding codon CAT of the first His in the catalytic motif of C$_4$ domain was mutated into Ala codon GCT; Km^R
E. coli GBred:: p15A-Ptet-vio-C4H2mut	*E. coli* GB05-red carring p15A-Ptet-vio-C4H1mut; the encoding codon CAC of the second His in the catalytic motif of C$_4$ domain was mutated into Ala codon GCC; Km^R
E. coli GBred:: p15A-Ptet-vio-C4H3mut	*E. coli* GB05-red carring p15A-Ptet-vio-C4H1mut; the encoding codon CAC of the third His in the catalytic motif of C$_4$ domain was mutated into Ala codon GCC; Km^R
M. xanthus DK1622 (*ΔmchA*)	Heterologous host for vioprolide expression; *Myxococcus xanthus* DK1622; *ΔmchA*, the myxochromide A gene cluster was deleted; Tet^R
M. xanthus::Ptet-violig	*M. xanthus* with *vio* gene cluster *Tn5-km^R-gent^R-tet^R-Ptet-violig* integrated into chromosome; *vioD* repaired, *vioA* and *vioZ* mutated; Tet^R, Km^R
M. xanthus::Ptet-vio	*M. xanthus* with *vio* gene cluster *Tn5-km^R-gent^R-tet^R-Ptet-vio* integrated into chromosome; *vioZ, vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptet* promoter; Tet^R, Km^R
M. xanthus::Ptn5-vio	*M. xanthus* with *vio* gene cluster *Tn5-km^R-vio* integrated into chromosome; *vioZ, vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptn5* promoter; Tet^R, Km^R
M. xanthus::Ptet-vio-int	*M. xanthus* with *vio* gene cluster *Tn5-km^R-gent^R-tet^R-Ptet-vio* integrated at *tetR* site of the chromosome; *vioZ, vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptet* promoter; Tet^R, Km^R
M. xanthus::Ptn5-vio-int	*M. xanthus* with *vio* gene cluster *Tn5-km^R-vio* integrated at *tetR* site of the chromosome; *vioZ, vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptn5* promoter; Tet^R, Km^R
M. xanthus::vio-orfNDel	*M. xanthus* with *Tn5-km^R-gent^R-tet^R-Ptet-vioArecov-vio-orfNDel* (N= 3, 4, 5, 6, 5&6, 7, 8, 9 or 3~9) integrated into chromosome; the *orf3, orf4, orf5, orf6, orf5&6, orf7, orf8, orf9 or orf3~9* in the *vio* gene cluster was deleted; *vioZ* mutated, *vioA* and *vioD* were repaired; Tet^R, Km^R
M. xanthus::Ptet-vioArecov-A1E331D	The *vio* gene cluster with mutated position 331 in A$_1$ domain was integrated into chromosome; *vioZ* mutated, *vioA* and *vioD* were repaired; Tet^R, Km^R
M. xanthus::Ptet-vioArecov-NMTmut	The *vio* gene cluster with the inactivated NMT domain was integrated into chromosome; *vioZ* mutated, *vioA* and *vioD* were repaired; Tet^R, Km^R
M. xanthus::Ptet-vioArecov-A1E331D-NMTmut	The *vio* gene cluster with mutated position 331 in A$_1$ domain and inactivated NMT domain was integrated into chromosome; *vioZ* mutated, *vioA* and *vioD* were repaired; Tet^R, Km^R
M. xanthus::Ptet-vio-FACdel	The *vio* gene cluster with the *faal-acp-c* deletion was integrated into chromosome; *vioZ, vioA* and *vioD* were repaired;Tet^R, Km^R
M. xanthus::Ptet-vio-C4H1mut	The *vio* gene cluster with the first His mutated in the catalytic motif of C$_4$ domain was integrated into chromosome; *vioZ, vioA* and *vioD* were repaired;Tet^R, Km^R
M. xanthus::Ptet-vio-C4H2mut	The *vio* gene cluster with the second His mutated in the catalytic motif of C$_4$ domain was integrated into chromosome; *vioZ, vioA* and *vioD* were repaired;Tet^R, Km^R
M. xanthus::Ptet-vio-C4H3mut	The *vio* gene cluster with the third mutated His in the catalytic motif of C$_4$ domain was integrated into chromosome; *vioZ, vioA* and *vioD* were repaired;Tet^R, Km^R
Burkholderia DSM7029	Heterologous host for vioprolide expression; *Burkholderia* K481-B101 (ATCC 53080; DSM 7029)
Burkholderia DSM7029::Ptet-vio	*Burkholderia* DSM7029 with *vio* gene cluster *Tn5-km^R-gent^R-tet^R-Ptet-vio* integrated into chromosome; *vioZ, vioA* and *vioD* were repaired;*vio* gene cluster was controlled by *Ptet* promoter; Km^R
Burkholderia DSM7029::Ptn5-vio	*Burkholderia* DSM7029 with *vio* gene cluster *Tn5-km^R-vio* integrated into chromosome; *vioZ, vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptn5* promoter; Km^R
Burkholderia DSM7029::vio-orfNDel	*Burkholderia* DSM7029 with *Tn5-km^R-gent^R-tet^R-Ptet-vioArecov-vio-orfNDel* (N= 3, 4, 5, 6, 5&6, 7, 8, 9 or 3~9) integrated into chromosome; the *orf3, orf4, orf5, orf6, orf5&6, orf7, orf8, orf9 or orf3~9* in the *vio* gene cluster was deleted; *vioZ* mutated, *vioA* and *vioD* were repaired; Km^R
Pseudomonas putida KT2440	Heterologous host for vioprolide expression; Amp^R

P. putida KT2440::Ptet-vio	*P.* putida KT2440 with *vio* gene cluster *Tn5-km^R-gent^R-tet^R-Ptet-vio* integrated into chromosome; *vioZ, vioA* and *vioD* were repaired;*vio* gene cluster was controlled by *Ptet* promoter; Amp^R, Km^R
P. putida KT2440::Ptn5-vio	*P.* putida KT2440 with *vio* gene cluster *Tn5-km^R-vio* integrated into chromosome; *vioZ, vioA* and *vioD* were repaired; *vio* gene cluster was controlled by *Ptn5* promoter; Amp^R, Km^R

Primers

ET21-cyclo-up	CCTCGTGTGTCCGGACTAC
Vio-5	GCCGCCTTCGAGCTCTAC
vio-cyclo-up	CTCGAGTTGTTCCTCGAGTACG
vio-cyclo-down	CGTAGAGACAGCCGAAAGAAGG
Vio9-Cfor	CGCCTGTGGTTCCTGCATCA
Vio9-Crev	AGGGTCATGTACAGCGTGGC
Vio-9	AGTGGCTCGAGAGTGCTTGTGACGAG
Vio-10	AGGAGCTCATCGAGCAGATCGGCATC
Vio-13	AGATGGCTCGACAGCTGCGTG
Vio-14	GATGATCGCGACGGACGTTTCC
V-zeo	AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTAG CTAGCCTTAAGAGCACGTGTTGAC
V-overlap-3	GGAGTCCTACTCAGAGTCTCAGTCCTGCTCCTCGGC
V-overlap-5	GCCGAGGAGCAGGACTGAGACTCTGAGTAGGACTCC
V-amp-5	GAGCGCCGCCGCCTTCTCGAGGAGAAGCGCAAGAACAGGAGGCAGCCGT GACATATGCTAGCTGACCTGT
V-amp-3	TGGCAGGTTGGGCGTCGCTTGGTCGGTCATTTCGAACCCCAGAGTCCCGCC TACCAATGCTTAATCAGTG
VioD-15A	CCTCGTCCGCCGCGTCCTTCAGCACCTCCTCGAATGCCGCCTGGGCCTCCG GGAACCGGCACAGCTCGAACAGGGCAATGGCGCTAGCGGAGTGTATACTG
vioD-neo	GACTCCTCGAGGAGGTACTGGAGCAGCTCGAGATCGTCCCCGGTGAGCGA TGCCGCAGAATCCCTCTCGTGCTTCATATCTCGCCTCCTTAAGCCTAGGTC AGAAGAACTCGTCAAGAAG
Vio-Tps5	TCGAATGCCGCCTGGGCCTCCGGGAACCGGCACAGCTCGAACAGGGCAAT GGCGAACCTCATTCCCTCATGATAC
P15AUpSeq	AGCACTTCACTGACACCCTCATC
genTetVi5	TTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGAAGG CACGAACCCAGTTGAC
genTetVi3	TCAACCGCTCGATGAGCTGGTGCATTGAGCCGGTCGACACCGACATGAGT GCCTCTTCTCTATCACTGATAGG
vioCDlig15	ACGTTCCATGTGGAGCTGCCAC
vioCDlig13	CGCAGAATCCCTCTCGTGCTTCACGGCTGCCTCCTGTTCTT
vioCDlig25	AAGAACAGGAGGCAGCCGTGAAGCACGAGAGGGATTCTGCG
vioCDlig23	TCATGTGATAGGCGGCGCTCTC
amp-ccdB5	CGCCGCCGCCTTCTCGAGGAGAAGCGCAAGAACAGGAGGCAGCCGTGATT TGTTTATTTTTCTAAATAC
amp-ccdB3	AGCGATGCCGCAGAATCCCTCTCGTGCTTCATATCTCGCCTCCTTAAGCCA GCCCCATACGATATAAGTTG
XmaJIhomolog-cm5	AGTGAGCAGGAGGCTCCTCGCACGGCGCTCGAAGCGGAGCTGGCCCGTAT CCTAGGGAGGAGCGCAGTCACCTCTAGATG
Homolog-cmccdb3	CGCCCAGGTCGAAGAACCGGTCATCACGGCCGACACGCTCCAGGCCGAGC AAGCCCGCTCATTAGGCGGGCTG
vioAseq5	TATGTGGTGCCACACGCCGGAGGAC
vioAseq32	CGTCACCGGCCTGACGTGGCACCAGG
orf3Del-cm5	GAACCACGGTTACGAAATCATCACGGTCGAGGTCCCGACCACGGAGCACC TAGGGAGGAGCGCAGTCACCTCTAGATG
orf3Del-cm3	TCACTCCGGCTGGTAGACCTGCGTGCGAGCCATCAGCTCGCGAGAGAAAC GCCTAGGTGGTTATGTGTGGGAGGGCTAACC
orf4Del-cm5	ATGGATGAATTCAACCGAATCTGGTCCGGGCTCGGCGCTTCCGTGCGCAA GCCTAGGGAGGAGCGCAGTCACCTCTAGATG
orf4Del-cm3	TCAGGAGCTCTGCGCCGGCTCGCTCCAGGGGAGGTGGCCCTCGGGAGTTC CCCTAGGTGGTTATGTGTGGGAGGGCTAACC
orf5Del-cm5	AGTCCCACCTGCGCCGCTGTTCCACGCGGCGCGCATTGGCAGGCCATGACC TAGGGAGGAGCGCAGTCACCTCTAGATG
orf5Del-cm3	CCCTGACCCTGCGCATAGAAGTCGAGCGTGCTCATGCCTTGCTCCCGGCCC TAGGTGGTTATGTGTGGGAGGGCTAACC
orf6Del-cm5	GCACGCTCGACTTCTATGCGCAGGGTCAGGGCGACCGGCTCATCGACCCCT

105

	AGGGAGGAGCGCAGTCACCTCTAGATG
orf6Del-cm3	GATTCCCAGCGGCTTCCGGCCGCTGGCCCTTCACTGCACCGATCGACGCTC
	CTAGGTGGTTATGTGTGGGAGGGCTAACC
orf7Del-cm5	GCGGCCGGAAGCCGCTGGGAATCAACAGCAAGGAGAACGTGCTTATGGCG
	CCTAGGGAGGAGCGCAGTCACCTCTAGATG
orf7Del-cm3	CGTAAGCAAACAGCGGGTGTGTGGAATCAGAGCTTGTTCAGGATCTTCCG
	CCTAGGTGGTTATGTGTGGGAGGGCTAACC
orf8Del-cm5	CCACACACCCGCTGTTTGCTTACGGAGAACTGAACATGGAAACACCCAAC
	CCTAGGGAGGAGCGCAGTCACCTCTAGATG
orf8Del-cm3	TCCTCCCTGAACCGGGAGGCCGTCCTCGTCGTGGCTCAGTACCGCCAGCCC
	CTAGGTGGTTATGTGTGGGAGGGCTAACC
orf9Delcm5	GGCCTCATCCAGTCCCTCAGGGGTGCACGAGGAGCATCAGGAGCAGTCAC
	CTAGGGAGGAGCGCAGTCACCTCTAGATG
orf9Delcm3	CATTGCCCTGTTCGAGCTGTGCCGGTTCCCGGAGGCCCAGGCGGCATTCGC
	CTAGGTGGTTATGTGTGGGAGGGCTAACC
orf3Delcheck5	GGACATGAATGCGGGAGTTATG
orf4Delcheck5	TCGCGAGCTGATGGCTCGCAC
orf34Delcheck3	GAGGTTCTGCTGCTCGATGAAC
orf56Delcheck51	TCGTCAACACGCTGGGAACTCC
orf56Delcheck52	CAGCTCTTCGAACAGCTCCAGC
orf56Delcheck3	GCGCCATAAGCACGTTCTCCTTG
orf78Delcheck5	ATGTTCGAGCCCTGGCGCCCTTG
orf78Delcheck3	GCTCCTGATGCTCCTCGTGCAC
orf9Delcheck3	CATTGCCCTGTTCGAGCTGTGCC
Ptetviozcmccdb5	TTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAGAGGCACTCC
	TAGGGAGGAGCGCAGTCACCTCTAGATG
Ptetviozcmccdb3	GCGGGTCGTAATGGAGCAGGCTCTGCATCAACCGCTCGATGAGCTGGTGC
	CTAGGAAGCCCGCTCATTAGGCGGGCT
Ptetviozrecov5	TTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAGAGGCACTCA
	TGGAGACGCTGCTCATTTCGG
vioZrecov3	CCTTTTGCGGACCTCGATGGACC
vioZampccdB5	CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGATT
	TGTTTATTTTTCTAAATAC
vioZampcounter3	GCGGGTCGTAATGGAGCAGGCTCTGCATCAACCGCTCGATGAGCTGGTGC
	AGCCCCATACGATATAAGTTG
vioZrecov5	CCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGACT
	TAAGGAGGTACAGATTATGGAGACGCTGCTCATTTCGG
VioA1mutcmccdB5	CTCGACGTGCCGCTGTACAACCTATACGGCCCGACAGAGGCGGCGATCGA
	CCTAGGGAGGAGCGCAGTCACCTCTAGATG
VioA1mutcmccdB3	CGGGAGGCAGGGGCGCGAGTTCCCGATCGAGCACGTGGAGTTGCGTGTTG
	CCTAGGAAGCCCGCTCATTAGGCGGGCT
VioA1E-D5	CTCGACGTGCCGCTGTACAACCTATACGGCCCGACAGAGGCGGCGATCGA
	CGTCAGCGCCTGGCAGTGC
VioA1E-D3	TCTCCAGCGATACCGGGAGGCAG
NMTmutcmccdB5	TGGCTCACATCCTCTCGCTCGCGCCGGAGCGGGTGTTGGAGATCGGGTGTC
	CTAGGGAGGAGCGCAGTCACCTCTAGATG
NMTmutcmccdB3	AGGATGATCGCGTCGTAGCTGCCCGGTTCGAACCCGCTGACGTCATCCGCC
	TAGGAAGCCCGCTCATTAGGCGGGCT
NMT-G-R-5	TGGCTCACATCCTCTCGCTCGCGCCGGAGCGGGTGTTGGAGATCGGGTGTC
	GAACGGGCCTCCTGCTGC
NMT-G-R-3	CACCGAGTTCAGGATGATCGCGTC
Module1-NheI-5	GCTAGCTAGCATGACTGCTTCCGTGAGTATCAGCC
Module1-NotI-3	ATAAGAATGCGGCCGCGATTGCCCCCTGCTCGGCGTGGATG
FA-AatII-5	GGAGGACGTCACCGCGCAGCATG
FA-NotI3	AGTGCGGCCGCCTACCCACCAGGCCCGGCGGAGGGCTTC
HisACP-NcoI5	ATACCATGGGCCATCATCATCATCATCACGACGCGGGCGGCTCCAGGGCG
FA-NotI3	AGTGCGGCCGCCTACCCACCAGGCCCGGCGGAGGGCTTC
KP-NheI5	CCTAGCTAGCATGTTCCAGGAGCGTGAGGGCCTGAC
KP-NotI3	AGTGCGGCCGCCTAGATTGCCCCCTGCTCGGCGTGGATG
VC4-cm5	TCGAGGAGGATCGCAAGCGTGGCTTCGACTTCTCGTCTGCGCCCCTGATGC
	CTAGGAAATAAATCCTGGTGTCCCTGTTGA
VC4-ccdB3	CTCCTCGAAGACAGGCTCACCGTCGATACGCTCGGAGAGCGCACGATCCA
	CCTAGGCAAAAAAAGCCCGCTCATTAGGCG
VC4I-5	TCGAGCGCTTCCTCGAGGAGGATC
VC4II-5	CCTGCTCCTGGACGGCTGGAG
VC4II-3	TGAGCCGCTGCTCCTCGAAGAC

106

VC4H1I-3	CTCCAGCCGTCCAGGAGCAGGTGGTGAGCACTCCAGAC
VC4H2I-3	CTCCAGCCGTCCAGGAGCAGGTGGGCATGACTCCAGAC
VC4H3I-3	CTCCAGCCGTCCAGGAGCAGGGCGTGATGACTCCAGAC

References

1 Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, *7*, 78. DOI: 10.1186/1471-2148-7-78.

2 Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Müller, R.; Stewart, A. F.; Zhang, Y. (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature Biotechnology*, **30** (5), 440–446.

3 Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, F. A.; Müller, R. (2005) Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chemistry & Biology*, **12**(3), 349–356. DOI: 10.1016/j.chembiol.2004.12.012.

4 Fu, J.; Wenzel, S. C.; Perlova, O.; Wang, J.; Gross, F.; Tang, Z.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. (2008) Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Research*, **36**(17), e113. DOI: 10.1093/nar/gkn499.

5 Wang, H.; Bian, X.; Xia, L.; Ding, X.; Müller, R.; Zhang, Y.; Fu, J.; Stewart, A. F. (2014) Improved seamless mutagenesis by recombineering using *ccdB* for counterselection. *Nucleic Acids Research*, **42**(5), e37. DOI: 10.1093/nar/gkt1339.

6 Zhang, Y.; Muyrers, J. P. P.; Testa, G.; Stewart, A. F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nature Biotechnology*, **18**, 1314–1317.

Manuscript

Biosynthesis and heterologous expression of vioprolide gene cluster revealing a C domain catalyzed glycerate esterification and a post-assembly maturation
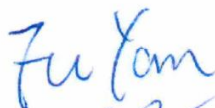
Fu Yan‡, David Auerbach‡, Yi Chai, Lena Keller, Qiang Tu, Youming Zhang*, Rolf Müller*

Author Contributions

Fu Yan designed the research, prepared the manuscript; performed sequence analysis, genetic engineering, compounds purification, designed and performed *in vitro* biochemical experiments. David Auerbach prepared the manuscript, purified compounds and elucidated chemical structures, designed and performed *in vitro* biochemical experiments. Yi Chai performed sequence analysis, cloned the *vio* gene cluster and performed genetic engineering. Lena Keller elucidated structures. Qiang Tu performed genetic engineering. Youming Zhang designed and supervised the research. Rolf Müller designed and supervised the research.
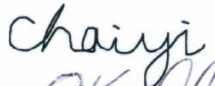
Signatures:

Fu Yan:

David Auerbach:

Yi Chai:

Lena Keller:

Qiang Tu:

Youming Zhang:

Rolf Müller:

3. DISCUSSION

3.1 General scope of this work

Natural products from myxobacteria are an abundant resource in the development of anti-infective and anticancer pharmaceuticals. The present thesis dealt with identification, elucidation and engineering of the biosynthetic pathway of vioprolides from *C. violaceus* Cb vi35. Heterologous expression was applied to validate the biosynthetic pathway and improve the production yield. Site-directed mutagenesis was carried out to expand the structural diversity. Biochemical investigations were performed to characterize the unusual initiation module responsible for the formation of previoprolides.

3.2 Insights into vioprolide biosynthesis

A big part of microbial natural products originate from PKS and (or) NRPS gene clusters. The genetic information of microorganisms became easily accessible in the recent years due to the development of next generation DNA sequencing technologies [1, 2]. Natural product biosynthetic gene clusters are to some extent predictable from microbial genomes with the aid of several bioinformatics tools such as antiSMASH [3], CLUSEAN [4], ClustScan [5], SBSPKS [6] and NRPSpredictor [7]. In this study, the genome sequence of *C. violaceus* Cb vi35 was obtained by Illumina MiSeq sequencing. *In silico* analysis of the Cb vi35 genome by antiSMASH identified a putative NRPS pathway producing a compound with a similar skeleton to vioprolide. The pathway is composed of two unusual modules and eight NRPS modules. Ten elongation steps on the NRPS machinery lead to the assembly of previoprolides, the precursors of vioprolides.

3.2.1 Assembly initiation

The assembly of vioprolides starts with the activation of long chain fatty acids. The lipid side chain is recruited from the fatty acid pool by the starting fatty acyl-AMP ligase (FAAL) domain of which the function was verified by *in vitro* biochemical investigations. FAALs are a new fatty acid activating enzyme family. They are firstly reported in *Mycobacterium tuberculosis* (Mtb) to activate and transfer long chain fatty acids to the carrier proteins of multifunctional PKSs [8]. Similar domains are also identified in other biosynthetic pathways such as mycosubtilin [9], alkylresorcylic acid [10] and telomycin [11]. FAALs are cognate proteins of fatty acyl-CoA ligase (FACL) with tiny difference in structure but play distinct roles. FACLs mainly involve in primary metabolism in Mtb, whereas FAALs take part in the biosynthesis of secondary metabolites such as phthiocerol dimycocerosates (PDIMs), sulfolipids, mycolic acids and mycobactin [12, 13]. The crystal structures of FACL and FAAL from Mtb reveal a large N-terminal domain and a small C-terminal domain [12–14]. The N-terminal domain contributes to AMP binding and fatty acid activation, whereas the C-terminal domain includes a CoA binding pocket and catalyzes fatty acyl-CoA ligation. In FACL, the C-terminal domain moves towards the N-terminal to ligate the activated acyl chain with CoA. A strong hydrophobic insertion motif, however, blocks the crosstalk between the two subdomains in FAAL and leads to the failure of acyl-CoA formation [13]. Indeed, sequence analysis revealed an insertion motif between the two subdomains of vioprolide FAAL (Fig. 3.1). The strong hydrophobic insertion loop between the two subdomains is more intuitive to be visualized from 3D-structural alignments (Fig. 3.2). The vioprolide FAAL is thus supposed to load fatty acyl moiety to the ACP domain without forming fatty acyl-CoA intermediates, which is also supported by the results from *in vitro* biochemical investigations.

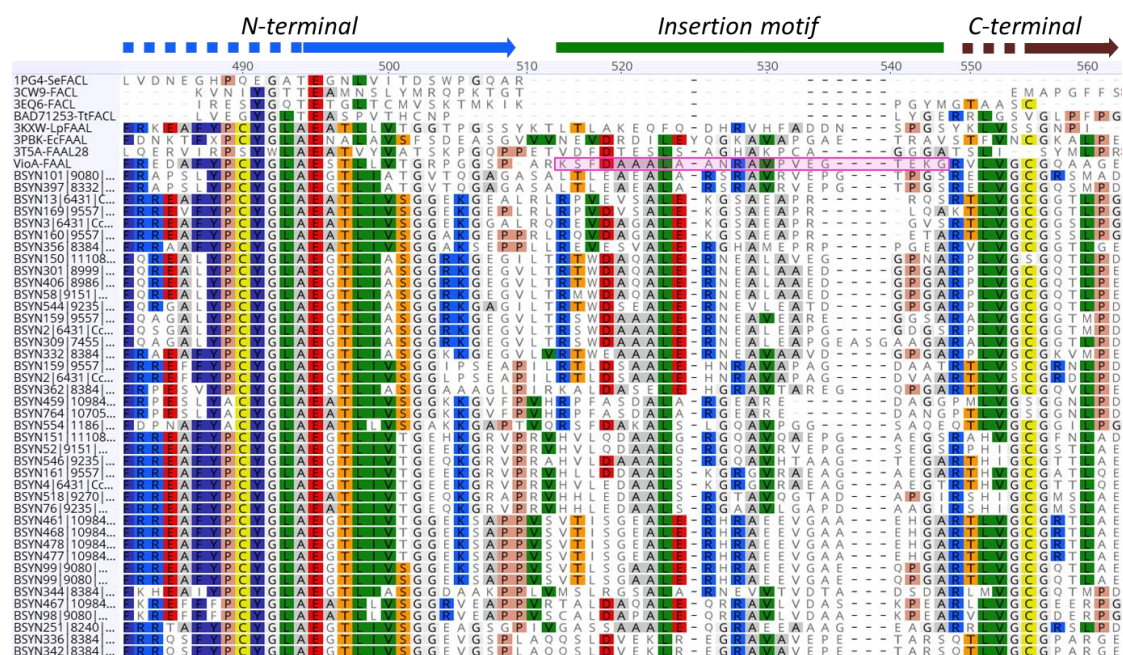**Fig. 3.1** Sequence alignment of myxobacterial FAALs with other reported FACLs and FAALs. The borders of N-terminal and C-terminal domains are indicated with arrows; insertion motifs are labeled with a green bar, and the insertion motif in vioprolide FAAL is highlighted in a pink box. Only part of 213 myxobacterial FAALs is shown here.
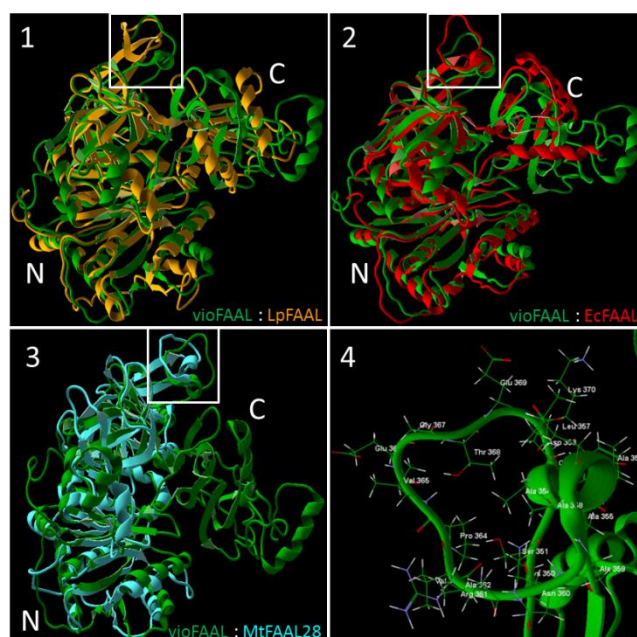

Fig. 3.2 Structure alignment of vioprolide FAAL (vioFAAL) with other reported FAALs. 1-3, alignment of vioFAAL with LpFAAL (*Legionella pneumophila* FAAL, PDB: 3KXW), EcFAAL (*E. coli* FAAL, PDB: 3PBK) and MtFAAL (*M. tuberculosis* FAAL, PDB: 3T5A), respectively; 4, enlarged insertion loop in vioFAAL. Strong hydrophobic insertion motifs are indicated with white boxes; N, N-terminal domain; C, C-terminal domain.

3.2.2 Glycerate incorporation and esterification

Glycerate is diverted to vioprolide assembly line by the FkbH domain in the module 2. FkbH-like proteins belong to haloacid dehalogenase (HAD) superfamily, which sequentially incorporate D-1, 3-biphosphoglycerate from the glycerol pool to form D-3-phosphoglyceryl-S-FkbH intermediates, then act as phosphatases to generate D-3-glyceryl-S-FkbHs, and finally transfer the D-glyceryl moiety to carrier proteins [15, 16]. FkbH-like proteins are also found to incorporate glyceryl units in the biosynthesis of several other natural products [15-24]. In vioprolide biosynthesis, the activated D-glycerate is the acceptor substrate of the C_1 domain which is, however, an ^LC_L C domain. Conversely, the D-glycerate is isomerized to L-glycerate which is the donor substrate of the ^DC_L C_2 domain. Similar situation was also discovered in cystomanamide biosynthetic machinery [23]. The ^DC_L C_5 domain of cystomanamide assembly line uses D-glycerate as the acceptor substrate which is the donor substrate of the ^LC_L C_6 domain. As glycerate is different from amino acids which are the common substrates of C domains, the configuration of glycerate probably has no effect on the condensation. Besides, the prediction of the C_1 and C_2 domain might be inaccurate due to the insufficient C domains with the same substrate specificity.

The fatty acyl-glycerate in previoprolide structure is similar to the acyl-glycerol in phospholipids. Glycerol acylation in the biosynthesis of phospholipids is catalyzed by glycerol-3-phosphate acyltransferase (GPAT) [25], whereas in vioprolide biosynthesis this process is catalyzed by a C domain. To the best of our knowledge, this is unique in secondary metabolism. To investigate the glycerate esterification function of the C_1 domain, we purified the related protein FAAL-ACP-C-FkbH-PCP from *E. coli*. The whole protein, however, was unstable and could not be detected by LC-MS so that the mass shift of the intermediates-attaching protein was indistinguishable from the intact protein. Cysteamine therefore was applied to release the intermediates mildly

3. Discusstion

from the carrier proteins. The successful detection of the released fatty acyl-glycerate by UPLC-HRMS confirmed the glycerate esterification function of the C_1 domain. Biosynthesis of natural products by PKS and NRPS machineries could involve dozens of reactions without generating free intermediates [26]. The mild release of carrier protein-bound intermediates with cysteamine provides a feasible way to investigate the individual reaction in biosynthetic pathways.

C domains play a critical role in the peptide chain extension in NRPs biosynthesis by forming a covalent bond between two carrier protein-bound substrates. Although C domains mostly catalyze amide bond formation between two amino acids, ester bond forming C domains (EBFCs) were also identified and characterized in the biosynthesis of fumonisin [27], C-1027 [28, 29] and cryptophycin [30]. According to the assembly lines and the final products, EBFCs are also supposedly involved in the biosynthesis of some other natural products such as enniatin [31, 32], FK520 [33] and antimycin [34]. Substrate recognition between these EBFCs could vary widely. For example, fumonisin EBFC (in FUM14) catalyzes the condensation between tricarboxylic acid and fumonisin backbone; C-1027 EBFC (SgcC5) condenses (*S*) -3-chloro-5-hydroxy-β-tyrosine with (*R*) -1-phenyl-1,2-ethanediol; and FK520 EBFC is supposed to catalyze intramolecular lactonization. Phylogenetic analysis showed no tight evolutionary relationship between these EBFCs, although several EBFCs catalyzing similar substrates could be classified into the same clade (Fig. S10). The catalytic motif of the EBFCs could be different from the consensus $HHXXXDX_{14}Y$ (Fig. S11). The second His-residue of the catalytic motif is usually critical for amide bond formation, and the Asp-residue is important for structural stabilization [35–37]. The Asp-residue is highly conserved in the EBFCs, whereas the second His-residue is changed to Leu in beauvericin EBFC and to Tyr in oxazolomycin EBFC (Fig. S11). It is unknown which residue takes

over the deprotonation function of the second histidine and catalyzes the formation of the ester bond. The blurry relationship between the protein sequence of EBFC and the distinctive function suggests that NRPS machinery is more diverse and complicated than expected. Mechanistic characterization of EBFCs should contribute to the generation of novel compounds by combinatorial biosynthesis in the future.

3.2.3 Proposed function of lipid side chain

The lipid chains of previoprolides were hydrolyzed after assembly. Precursor maturation was also discovered in the biosynthesis of xenocoumacin [38], pyoverdine [39, 40], colibactin [41] and naphthyridinomycin [42], the N-terminal fatty acid or aminoacyl chains of their precursors are excised after assembly by periplasmic proteases XcnG, PvdQ, ClbP and NapG, respectively. Caerulomycin A [43], quinocarcin [44] and telomycin [11] are also synthesized as precursor before being trimmed by the amidohydrolase CrmL, peptidase Qcn1 and acylase Tem25, respectively. In addition, biosynthesis of didemnin [45], saframycin [46] and nocardicin [47] involve similar precursor maturation process but the respective protease is unknown. Several proteases similar to XcnG, PvdQ, ClbP, CrmL, NapG or Tem25 were found in *C. violaceus* Cb vi35 and *M. xanthus* DK1622 genomes. The lipid chains of previoprolides may be hydrolyzed by one of these enzymes before extracellular secretion.

The activation of fatty acids is probably essential to vioprolide biosynthesis. Deletion of the FAAL-ACP-C tridomain abolished the production of vioprolides (Fig. 3.3), although the abolishment of vioprolide production may also result from incorrect protein folding of the shortened assembly line. The activation of fatty acids may provide a high potential energy to trigger the initiation of assembly line and drive the biosynthesis towards the C-terminus. In addition, the

fatty acids may control the regioselectivity of the TE domain in the biosynthesis of previoprolides. The TE-mediated macrocyclization takes place only between the -COOH group of the methyl-valine and the β-OH group, rather than the α-OH group, of the glycerate. Meanwhile, no cyclization was discovered between the -COOH group of the methyl-valine and the β-OH group of the serine during mutagenesis of hydroxyl-vioprolides. Similar instance was also found in the biosynthesis of calcium dependent antibiotics (CDA) in which the lipid chain length (C_6) is critical to determine the regioselectivity of TE domain [48, 49]. The acylation thus prevented the incorrect cyclization and ensured the final structure. In several natural products biosynthesis such as xenocoumacin and amicoumacin [50], acylation serves as a self-protection group so that the prodrug is nontoxic to the producer. However, the lipid side chain in previoprolide is more likely to serve as a guide for exportation, because no obvious change in cytotoxicity was observed in the acylated vioprolide B. The structure of the acylated glycerate in previoprolides simulates phospholipids – one of the major components of cell membrane. Membrane-guiding and extracellular secretion prevented the accumulation of vioprolides in cytosol and avoided the potential feedback inhibition on the expression of the *vio* gene cluster.
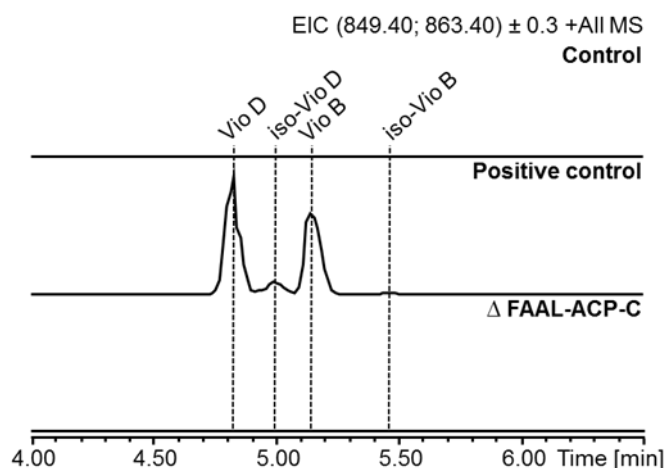
Fig. 3.3 HPLC-MS analysis of vioprolides from ΔFAAL-ACP-C mutants. The information of the muntant *M. xanthus*::Ptet-vio-FACdel (ΔFAAL-ACP-C) is shown in the strain list in supplementary. *M. xanthus* DK1622 (Δ*mchA*) was set as negative control, *M. xanthus*::Ptet-vioArecov-vio was set as positive control. EIC *m/z* for 849.4 [M+H]$^+$ (vioprolide D and isomer) and 863.4 [M+H]$^+$ (vioprolide B and isomer) are shown. Intensity is adjusted to the same range.

3.3 Heterologous expression of the vioprolide biosynthetic pathway

We initially attempted to validate the *vio* gene cluster in the native producer by gene knock-out. Genetic manipulation, however, was inaccessible in *C. violaceus* Cb vi35. Therefore, we tried to clone the putative *vio* gene cluster and express it in heterologous hosts. The cloning was not straightforward because of the large size of the *vio* gene cluster. Only a part of the gene cluster was obtained by screening the cosmid library, and the remaining parts were cloned by PCR and Red/ET recombination. Red/ET recombination is a homologous recombination-based genetic engineering method, which has an advantage to edit DNA sequence of interest precisely without limit of the targeting sites [51–53]. It has recently been developed as a simple and fast method to clone natural product biosynthetic gene clusters directly from genomic DNA [54–56]. Most of the modifications on the *vio* gene cluster were finished by Red/ET recombination.

Heterologous expression reveals great advantages in characterization of biosynthetic pathways. The *vio* gene cluster was validated by successful expression in heterologous hosts. The core gene cluster (*vioZ, vioA~D*) was also determined by means of gene knock-out combined with heterologous expression. Furthermore, the production titer of vioprolides increased significantly in heterologous hosts *M. xanthus* DK1622 (*ΔmchA*) and *Burkholderia* DSM7029. Especially in *M. xanthus*::Ptet-vio, the overall production of vioprolides increased around 8-fold comparing with the native producer. The heterologous hosts grow faster than the native producer and the regulation network on the *vio* gene cluster was eliminated by replacing the native promoter with constitutive *Ptet* or *Ptn5* promoter. The constitutive promoter ensured constant transcription of the *vio* gene cluster. In addition, the heterologous host *M. xanthus* DK1622 is evolutionarily close to *C. violaceus* Cb vi35. Both strains belong to *Cystobacteraceae* in the order *Myxococcales*. The GC content of the *vio* gene cluster (68.9%) is the same with the genome of *M. xanthus* DK1622. The *vio* gene cluster probably adapted well in *M. xanthus* DK1622 thus achieving a high expression level. Weak genetic adaptation or inefficient transcription of the *vio* gene cluster in *P. putida* and *E. coli* is probably the reason of the low and null production yield, promoter exchange or codon optimization may improve the vioprolide production.

3.4 Production improvement via bioprocess optimization

The successful expression of the *vio* gene cluster in *M. xanthus* is a good starting point to increase the production yield further by bioprocess optimization. Although the *vio* gene cluster was successfully expressed in *M. xanthus* DK1622 (*ΔmchA*) after chromosomal transposition, it is unclear whether the transposition disrupted endogenous genes and affected vioprolide production. The *vio* gene

cluster was therefore integrated at the *tetR* site in the chromosome of *M. xanthus* DK1622 (*ΔmchA*) by homologous recombination to avoid interrupting other native genes. Similar to transposition mutants, the *tetR* integration mutants *M. xanthus*::Ptet-vio-int and *M. xanthus*::Ptn5-vio-int also successfully produced vioprolides (Fig. 3.4). The *tetR* integration mutants were sent to our collaborator in HZI (Braunschweig) for bioprocess optimization. Their preliminary results from bioreactor revealed that the overall production of vioprolides increased to almost 1 g/L via bioprocess optimization on the mutant *M. xanthus*::Ptn5-vio-int (unpublished). Further optimization on fermentation conditions could probably achieve a higher production titer. Moreover, considering the *Ptet* promoter is more efficient than *Ptn5* promoter to drive the *vio* gene cluster in *M. xanthus*, the mutant *M. xanthus*::Ptet-vio-int has a potential to produce vioprolides in a higher yield.
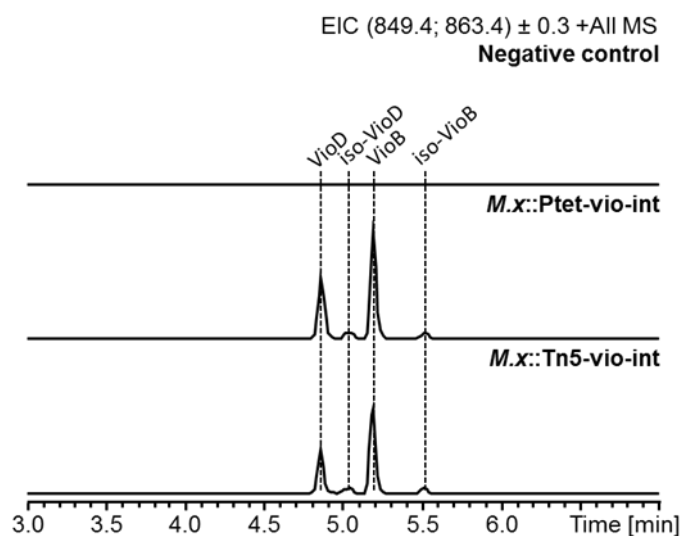

Fig. 3.4 HPLC-MS analysis of vioprolides produced by *M. xanthus*::Ptet-vio-int and *M. xanthus*::Ptn5-vio-int. The production of vioprolide B-D is indicated with dashed lines, extracted ion chromatogram (EIC) *m/z* for 849.4 [M+H]$^+$ (vioprolide D and isomer) and 863.4 [M+H]$^+$ (vioprolide B and isomer) is shown. *M. xanthus* DK1622 (*ΔmchA*) was set as negative control. Intensity is set to the same range.

3.5 Creation of molecular diversity by mutagenesis

So far four types of vioprolides have been reported from *C. violaceus* Cb vi35. In this work, we also identified linear Me-vioprolide B, linear vioprolides, previoprolides and linear previoprolides in the native and heterologous producers (Fig. 3.5). The linear previoprolides may be a result of direct release from the TE domain and (or) ring opening of the circular previoprolides, whereas linear vioprolides are probably hydrolyzed from linear previoprolides and (or) circular vioprolides.

NRPS machineries may show some substrate tolerance [57]. A-domains are critical for substrate selection and activation. Novel derivatives could thus be obtained by altering the specificity-conferring code of A domains, as represented in the engineering of luminmide [58], gramicidin S [59], hormaomycin [60], calcium dependent antibiotics (CDA) [61], just to name a few examples [62]. In this work, vioprolides B1~D1 were generated by site-directed mutagenesis in the A_1 domain; vioprolides B2~D2 were produced by inactivation of the NMT domain; and vioprolides B3~D3 were created by mutating both A_1 and NMT domains. Similar to the linear vioprolides produced by the native *vio* gene cluster, linear derivatives were also found in the site-mutated mutants (Fig. 3.5). The additional –OH group in the position 2 of vioprolides was achieved by altering the position 331 of the analine activating A_1 domain. Resulting incorporation of serine instead of analine in the molecule increased the hydrophilicity of vioprolide B1 and resulted in around 100-fold less cytotoxicity than vioprolide B. Although the mode of action of vioprolide B is unclear at present, the addition of the –OH group probably interferes with the target binding site thus reduces the cytotoxicity. The compounds identified and generated in this study are summarized in Fig. 3.5. The production of vioprolide derivatives displays the potential to create more novel structures by engineering

the NRPS machinery. The expanded chemical diversity provided a compound library for candidate drug screening in the future.
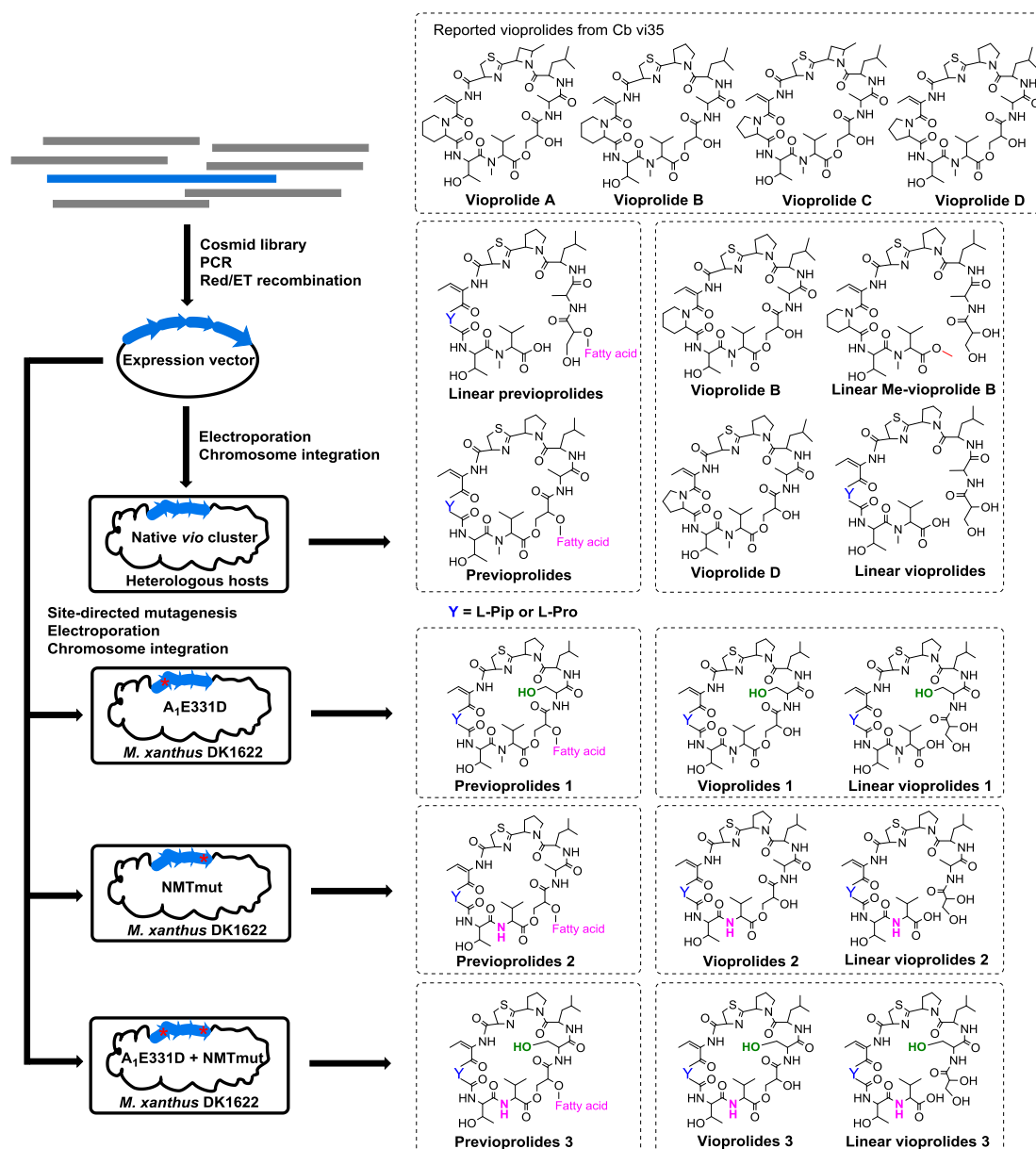

Fig. 3.5 Heterologous expression of *vio* gene cluster and the compounds generated in this study. Vioprolide B (6.6 mg), vioprolide D (15.7 mg), linear Me-vioprolide B (3 mg), vioprolide B1 (1.4 mg), vioprolide D1 (2.6 mg) and previoprolides (1.2 ~ 11.9 mg) were isolated from 10 L cultures of *M. xanthus* mutants, whereas others were limited to LC-MS and MS/MS characterization due to the low abundancy. Vioprolide D, linear Me-vioprolide B, vioprolide D1 and previoprolide B were validated by NMR.

3.6 Potential of lipopeptide biosynthesis in myxobacteria

FAAL-like proteins could serve as probe to screen candidate lipopeptide biosynthetic gene clusters in microbes, because their main function is activation of fatty acids in secondary metabolites biosynthesis. Alignment of vioprolide FAAL protein sequence in our in-house database displayed 244 myxobacterial proteins annotated as FACL. Detailed sequence analysis revealed that 213 of them are factually FAAL with distinctive insertion motifs (Fig. 3.1). Sixty eight of the 213 FAALs locate in NRPS or PKS/NRPS hybrid gene clusters (Fig. 3.6), whereas the others are discrete. Statistical analysis showed that *Corallococcus* and *Sorangium* species contain many FAAL-containing gene clusters (Fig. 3.6). The identified FAAL-containing gene clusters are listed in Table 3.1. Although FAAL-activated fatty acids may not exist in the final products, they could be involved in the biosynthesis of lipopeptide precursors. These findings suggested considerable undiscovered lipopeptide diversity in myxobacteria. Up to now, most of the identified lipopeptides are from *Bacillus*, *Pseudomonads* and *Streptomyces* etc. Only three classes of lipopeptides have been identified from myxobacteria (myxochromides, cystomanamides and previoprolides; cystomanamides and previoprolides are not lipopeptide in the strict definition). The potential of lipopeptide biosynthesis in myxobacteria is probably severely underestimated.
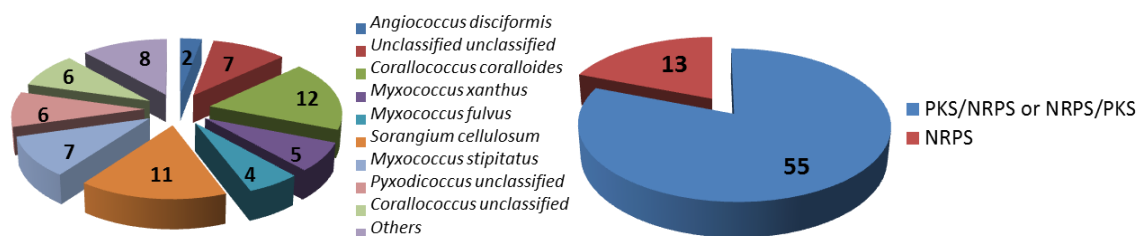

Fig. 3.6 Statistical analysis of the 68 FAAL-containing gene clusters in myxobacteria. The FAAL-containing gene clusters were retrieved from our in-house Mxbase. The number of the gene clusters and the respective species are shown in the left side; the type of of the 68 FAAL-containing gene cluster is shown in the right side.

Table 3.1 FAAL-containing gene clusters identified in our in-house database

Strain	FAAL-containing gene cluster*
Angiococcus disciformis AngGT8	BSYN309, BSYN315
Aetherobacter unclassified MSr9329	BSYN609
Aetherobacter unclassified MSr9335	BSYN666
Aetherobacter fasciculatus MSr9337	BSYN626
Corallococcus unclassified MCy10984	BSYN459, BSYN467, BSYN468, BSYN474, BSYN478, **Corramycin**
Corallococcus coralloides Ccc1071	BSYN5, BSYN3, BSYN13
Corallococcus coralloides MCy9080	BSYN98, BSYN101, BSYN99, BSYN98
Corallococcus coralloides ST201330	BSYN768, BSYN764
Cystobacter unclassified MCy9101	BSYN587
Cystobacter velatus Cbv34	BSYN381
Cystobacter fuscus MCy9118	**Cystomanamide**
Cystobacter violaceus Cb vi35	**Vioprolide**
Haliangium ochraceum MNa9133	BSYN63
Myxococcus xanthus DK897	BSYN406
Myxococcus xanthus MCy9151	BSYN52, BSYN50
Myxococcus xanthus MxA47	BSYN301, BSYN292
Myxococcus fulvus MCy11108	BSYN151, BSYN152
Myxococcus fulvus MCy9270	BSYN518, BSYN490
Myxococcus stipitatus MCy9235	BSYN78, BSYN76, BSYN546, BSYN71, BSYN544, BSYN75, BSYN547
Pyxodicoccus unclassified MCy9557	BSYN161, BSYN159, BSYN165, BSYN169, BSYN158, BSYN160
Sorangium cellulosum Soce26	BSYN229, BSYN244
Sorangium cellulosum Soce38	BSYN749
Sorangium cellulosum Soce56	BSYN258
Sorangium cellulosum Soce307	BSYN715
Sorangium cellulosum Soce836	BSYN283, BSYN266
Sorangium cellulosum Soce1128	BSYN206, BSYN194
Sorangium cellulosum MSr11115	BSYN191
Sorangium cellulosum SoceGT47	BSYN140
Stigmatella erecta Pde77	BSYN554
unclassified unclassified Ar3548	BSYN350, BSYN363, BSYN356, BSYN354, BSYN342, BSYN332, BSYN344
unclassified unclassified Ccc127	BSYN404, BSYN397

* only cystomanamide, corramycin (unpublished) and vioprolide (unpublished) gene clusters have been characterized; other gene clusters are given in original name.

3.7 The unsolved mystery − mechanism of Maz formation

As vioprolide A and C are not produced in *C. violaceus* Cb vi35 and heterologous hosts, it is not yet certain which protein catalyzed the formation of Maz. The next step is to find appropriate media and cultivating conditions to activate the production of vioprolide A and C. Only when vioprolide A and C are produced, could we find the precursor of Maz by isotope feeding. There are two reasons to consider methionine as the precursor of the 4-methylazetidine carboxylic acid (Maz). On the one hand, methyl groups are transferred from methionine by *S*-adenosylmethionine (SAM) dependent methyltransferase in numerous methylation reactions in the secondary metabolites biosynthesis [63]. The *vio* gene cluster does contain an *S*-adenosylmethionine (SAM) dependent methyltransferase (ORF6). On the other hand, the formation of azetidine carboxylic acid (AZE) has been shown to be relevant to the methionine cycle in gramineous plants [64–68]. Considering that some SAM-dependent methyltransferases could rearrange the carbon skeleton [63], the amino acids with similar carbon skeleton with Maz, such as proline and valine, could also be counted as candidate precursors. If the production of vioprolide A and C could be activated in *C. violaceus* Cb vi35, it is possible to find the Maz catalyzing protein by quatitative transcriptomics or quatitative proteomics strategy. We could cultivate *C. violaceus* Cb vi35 in different conditions and analyze the change of transcriptome or proteome.

3.8 Perspectives

The elucidation of vioprolide biosynthetic mechanism deepened our knowledge about natural product biosynthesis and established a platform to produce vioprolide analogues. The improvement of production yield sets the

stage for pharmacological and toxicological research as well as industrial production of vioprolides.

The research on the MOA of vioprolides would be a prospective direction of this project. If the cytotoxic and antifungal mechanisms of vioprolides could be elucidated in the future, more derivatives may be rationally generated to increase the specificity and activity to the target. High toxicity of vioprolides A-C on both healthy and carcinoma cells hindered their development to anti-cancer agents. It is thus necessary to reduce their toxic effects on healthy cells but retain the cytotoxicity towards cancer cells by structural modifications. Conjugating vioprolides with tumor specific monoclonal antibodies could also be an option to reduce their cytotoxicity on healthy cells. In addition, vioprolide D is a promising antifungal candidate. The research on its MOA is urgent because of the limited number of antifungal pharmaceuticals on the market and the severe situation of antifungal resistance. Moreover, vioprolides have potential to treat diseases susceptible to interferon α and β (US Patent: US 2010/0028298 A1), the immune-enhancing mechanism of vioprolides will also be of future interest. The investigations on these topics will proceed faster in the future due to the increased and stable availability of vioprolides.

The elucidation of Maz forming mechanism would be valuable for both fundamental research and industrial application. The Maz catalytic protein could be an element for combinatorial biosynthesis. Novel structures may be generated by introducing Maz catalytic protein into other L-Pro incorporating megasynthetases (e.g. myxochromide NRPS). Although we were not able to elucidate the mechanism for Maz formation in this study, activating the production of vioprolide A and C in *C. violaceus* Cb vi35 should lead to the identification of the precursor and the protein involve in Maz formation. As Maz is not produced in heterologous systems, it is possible to generate a mutant producing only vioprolide D by deleting the *vioZ* gene, which may elevate the

production yield of vioprolide D and could save considerable effort for downstream purifications.

3.9 References

1   van Vliet, Arnoud H M (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiology Letters*, **302** (1), 1–7.

2   Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, **10** (6), 563–569.

3   Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H. (2015) antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 43(W1):W237-243.

4   Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D.H., Wohlleben, W. (2009) CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology*, **140** (1-2), 13–17.

5   Starcevic, A., Zucko, J., Simunkovic, J., Long, P.F., Cullum, J., Hranueli, D. (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Research*, **36** (21), 6882–6892.

6   Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., Mohanty, D. (2010) SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Research*, **38** (Web Server issue), W487-96.

7   Rottig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., Kohlbacher, O. (2011) NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, **39** (Web Server issue), W362-7.

8   Trivedi, O.A., Arora, P., Sridharan, V., Tickoo, R., Mohanty, D., Gokhale, R.S. (2004) Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature*, **428** (6981), 441–445.

9   Hansen, D.B., Bumpus, S.B., Aron, Z.D., Kelleher, N.L., Walsh, C.T. (2007) The loading module of mycosubtilin: an adenylation domain with fatty acid selectivity. *Journal of the American Chemical Society*, **129** (20), 6366–6367.

10 Hayashi, T., Kitamura, Y., Funa, N., Ohnishi, Y., Horinouchi, S. (2011) Fatty Acyl-AMP Ligase Involvement in the Production of Alkylresorcylic Acid by a *Myxococcus xanthus* Type III Polyketide Synthase. *ChemBioChem*, **12** (14), 2166–2176.

11 Fu, C., Keller, L., Bauer, A., Brönstrup, M., Froidbise, A., Hammann, P., Herrmann, J., Mondesert, G., Kurz, M., Schiell, M., Schummer, D., Toti, L., Wink, J., Müller, R. (2015) Biosynthetic Studies of Telomycin Reveal New Lipopeptides with Enhanced Activity. *Journal of the American Chemical Society*, **137** (24), 7692–7705.

12 Mohanty, D., Sankaranarayanan, R., Gokhale, R.S. (2011) Fatty acyl-AMP ligases and polyketide synthases are unique enzymes of lipid biosynthetic machinery in *Mycobacterium tuberculosis*. *Tuberculosis (Edinburgh, Scotland)*, **91** (5), 448–455.

13 Arora, P., Goyal, A., Natarajan, V.T., Rajakumara, E., Verma, P., Gupta, R., Yousuf, M., Trivedi, O.A., Mohanty, D., Tyagi, A., Sankaranarayanan, R., Gokhale, R.S. (2009) Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nature Chemical Biology*, **5** (3), 166–173.

14 Zhang, Z., Zhou, R., Sauder, J.M., Tonge, P.J., Burley, S.K., Swaminathan, S. (2011) Structural and functional studies of fatty acyl adenylate ligases from *E. coli* and L. pneumophila. *Journal of Molecular Biology*, **406** (2), 313–324.

15 Dorrestein, P.C., van Lanen, S.G., Li, W., Zhao, C., Deng, Z., Shen, B., Kelleher, N.L. (2006) The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. *Journal of the American Chemical Society*, **128** (32), 10386–10387.

16 Zhang, F., He, H.Y., Tang, M.C., Zhou, Q., Tang, G.L. (2011) Cloning and elucidation of the FR901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate as starter units. *Journal of the American Chemical Society*, **138** (8), 2452–2462.

17 Wu, K., Chung, L., Revill, W.P., Katz, L., Reeves, C.D. (2000) The FK520 gene cluster of *Streptomyces hygroscopicus var. ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene*, **251** (1), 81–90.

18 Zhao, C., Coughlin, J.M., Ju, J., Zhu, D., Wendt-Pienkowski, E., Zhou, X., Wang, Z., Shen, B., Deng, Z. (2010) Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an "acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *Journal of Biological Chemistry*, **285** (26), 20097–20108.

19 Sun, Y., Hong, H., Gillies, F., Spencer, J.B., Leadlay, P.F. (2008) Glyceryl-S-acyl carrier protein as an intermediate in the biosynthesis of tetronate antibiotics. *ChemBioChem : a European journal of chemical biology*, **9** (1), 150–156.

20 Sun, Y., Hahn, F., Demydchuk, Y., Chettle, J., Tosin, M., Osada, H., Leadlay, P.F. (2010) In vitro reconstruction of tetronate RK-682 biosynthesis. *Nature Chemical Biology*, **6** (2), 99–101.

21 Emmert, E.A.B., Klimowicz, A.K., Thomas, M.G., Handelsman, J. (2004) Genetics of Zwittermicin A production by *Bacillus cereus*. *Applied and Environmental Microbiology*, **70** (1), 104–113.

22 Kevany, B.M., Rasko, D.A., Thomas, M.G. (2009) Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. *Applied and Environmental Microbiology*, **75** (4), 1144–1155.

23 Etzbach, L., Plaza, A., Garcia, R., Baumann, S., Müller, R. (2014) Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Organic Letters*, **16** (9), 2414–2417.

24 Jahns, C., Hoffmann, T., Müller, S., Gerth, K., Washausen, P., Höfle, G., Reichenbach, H., Kalesse, M., Müller, R. (2012) Pellasoren: structure elucidation, biosynthesis, and total synthesis of a cytotoxic secondary metabolite from *Sorangium cellulosum. Angewandte Chemie (International ed. in English)*, **51** (21), 5239–5243.

25 Zhang, Y.-M. and Rock, C.O. (2008) Thematic review series: Glycerolipids. Acyltransferases in bacterial glycerophospholipid synthesis. *Journal of Lipid Research*, **49** (9), 1867–1874.

26 Belecki, K. and Townsend, C.A. (2013) Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *Journal of the American Chemical Society*, **135** (38), 14339–14348.

27 Zaleta-Rivera, K., Xu, C., Yu, F., Butchko, R.A.E., Proctor, R.H., Hidalgo-Lara, M.E., Raza, A., Dussault, P.H., Du, L. (2006) A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation

of tricarballylic esters in the biosynthesis of fumonisins. *Biochemistry*, **45** (8), 2561–2569.

28 Lin, S., van Lanen, S.G., Shen, B. (2009) A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (11), 4183–4188.

29 Lin, S., Huang, T., Horsman, G.P., Huang, S.-X., Guo, X., Shen, B. (2012) Specificity of the ester bond forming condensation enzyme SgcC5 in C-1027 biosynthesis. *Organic Letters*, **14** (9), 2300–2303.

30 Ding, Y., Rath, C.M., Bolduc, K.L., Hakansson, K., Sherman, D.H. (2011) Chemoenzymatic synthesis of cryptophycin anticancer agents by an ester bond-forming non-ribosomal peptide synthetase module. *Journal of the American Chemical Society*, **133** (37), 14492–14495.

31 Haese, A., Schubert, M., Herrmann, M., Zocher, R. (1993) Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. *Molecular Microbiology*, **7** (6), 905–914.

32 Hornbogen, T., Glinski, M., Zocher, R. (2000) Biosynthesis of the mycotoxin enniatin in *Fusarium species*. *Mycotoxin Research*, **16 Suppl 1**, 84–87.

33 Gatto, G.J., JR, McLoughlin, S.M., Kelleher, N.L., Walsh, C.T. (2005) Elucidating the substrate specificity and condensation domain activity of FkbP, the FK520 pipecolate-incorporating enzyme. *Biochemistry*, **44** (16), 5993–6002.

34 Yan, Y., Zhang, L., Ito, T., Qu, X., Asakawa, Y., Awakawa, T., Abe, I., Liu, W. (2012) Biosynthetic pathway for high structural diversity of a common dilactone core in antimycin production. *Organic Letters*, **14** (16), 4142–4145.

35 Keating, T.A., Marshall, C.G., Walsh, C.T., Keating, A.E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, **9** (7), 522–526.

36 Bergendahl, V., Linne, U., Marahiel, M.A. (2002) Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry / FEBS*, **269** (2), 620–629.

37 Roche, E.D. and Walsh, C.T. (2003) Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry*, **42** (5), 1334–1344.

38 Reimer, D., Pos, K.M., Thines, M., Grün, P., Bode, H.B. (2011) A natural prodrug activation mechanism in nonribosomal peptide synthesis. *Nature Chemical Biology*, **7** (12), 888–890.

39 Drake, E.J. and Gulick, A.M. (2011) Structural characterization and high-throughput screening of inhibitors of PvdQ, an NTN hydrolase involved in pyoverdine synthesis. *ACS Chemical Biology*, **6** (11), 1277–1286.

40 Schalk, I.J. and Guillon, L. (2013) Pyoverdine biosynthesis and secretion in *Pseudomonas aeruginosa*: implications for metal homeostasis. *Environmental Microbiology*, **15** (6), 1661–1673.

41 Brotherton, C.A. and Balskus, E.P. (2013) A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *Journal of the American Chemical Society*, **135** (9), 3359–3362.

42 Pu, J.-Y., Peng, C., Tang, M.-C., Zhang, Y., Guo, J.-P., Song, L.-Q., Hua, Q., Tang, G.-L. (2013) Naphthyridinomycin biosynthesis revealing the use of leader peptide to guide nonribosomal peptide assembly. *Organic Letters*, **15** (14), 3674–3677.

43 Zhu, Y., Fu, P., Lin, Q., Zhang, G., Zhang, H., Li, S., Ju, J., Zhu, W., Zhang, C. (2012) Identification of caerulomycin A gene cluster implicates a tailoring amidohydrolase. *Organic Letters*, **14** (11), 2666–2669.

44 Hiratsuka, T., Koketsu, K., Minami, A., Kaneko, S., Yamazaki, C., Watanabe, K., Oguri, H., Oikawa, H. (2013) Core assembly mechanism of quinocarcin/SF-1739: bimodular complex nonribosomal peptide synthetases for sequential mannich-type reactions. *Chemistry & Biology*, **20** (12), 1523–1535.

45 Xu, Y., Kersten, R.D., Nam, S.-J., Lu, L., Al-Suwailem, A.M., Zheng, H., Fenical, W., Dorrestein, P.C., Moore, B.S., Qian, P.-Y. (2012) Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *Journal of the American Chemical Society*, **134** (20), 8625–8632.

46 Koketsu, K., Watanabe, K., Suda, H., Oguri, H., Oikawa, H. (2010) Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms. *Nature Chemical Biology*, **6** (6), 408–410.

47 Davidsen, J.M., Bartley, D.M., Townsend, C.A. (2013) Non-ribosomal propeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI. *Journal of the American Chemical Society*, **135** (5), 1749–1759.

48 Grunewald, J. and Marahiel, M.A. (2006) Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. *Microbiology and Molecular Biology Reviews*, **70** (1), 121–146.

49 Grünewald, J., Sieber, S.A., Marahiel, M.A. (2004) Chemo- and regioselective peptide cyclization triggered by the N-terminal fatty acid chain length: the recombinant cyclase of the calcium-dependent antibiotic from *Streptomyces coelicolor*. *Biochemistry*, **43** (10), 2915–2925.

50 Li Y., Xu Y., Liu L., Han Z., Lai P.Y., Guo X., Zhang X., Lin W., Qian P.Y. (2012) Five new amicoumacins isolated from a marine-derived bacterium *Bacillus subtilis*. Marine Drugs, **10**(2):319-328.

51 Muyrers, J.P., Zhang, Y., Buchholz, F., Stewart, A.F. (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes & development*, **14** (15), 1971–1982.

52 Zhang, Y., Muyrers, J.P.P., Rientjes, J., Stewart, A.F. (2003) Phage annealing proteins promote oligonucleotide-directed mutagenesis in *Escherichia coli* and mouse ES cells. *BMC Molecular Biology*, **4** (1), 1.

53 Muyrers, J.P.P., Zhang, Y., Benes, V., Testa, G., Rientjes, J.M.J., Stewart, A.F. (2004) ET recombination: DNA engineering using homologous recombination in E. coli. *Methods in Molecular Biology (Clifton, N.J.)*, **256**, 107–121.

54 Yin, J., Hoffmann, M., Bian, X., Tu, Q., Yan, F., Xia, L., Ding, X., Stewart, A.F., Müller, R., Fu, J., Zhang, Y. (2015) Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Scientific Reports*, **5**, 15081.

55 Liu, Q., Shen, Q., Bian, X., Chen, H., Fu, J., Wang, H., Lei, P., Guo, Z., Chen, W., Li, D., Zhang, Y. (2016) Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in *Bacillus subtilis* via Red/ET recombineering. *Scientific Reports*, **6**, 34623.

56 Wang, H., Li, Z., Jia, R., Hou, Y., Yin, J., Bian, X., Li, A., Müller, R., Stewart, A.F., Fu, J., Zhang, Y. (2016) RecET direct cloning and Redalphabeta recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nature Protocols*, **11** (7), 1175–1190.

57 Calcott M.J., Ackerley D.F. (2014) Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. Biotechnology Letters, **36**(12):2407-2416.

58 Bian, X., Plaza, A., Yan, F., Zhang, Y., Müller, R. (2015) Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives in vivo. *Biotechnology and Bioengineering*, **112** (7), 1343–1353.

134

59 Kries, H., Wachtel, R., Pabst, A., Wanner, B., Niquille, D., Hilvert, D. (2014) Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angewandte Chemie (International ed. in English)*, **53** (38), 10105–10108.

60 Crusemann, M., Kohlhaas, C., Piel, J. (2013) Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chemical Science*, **4** (3), 1041–1045.

61 Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A.W., Smith, C.P., Micklefield, J. (2012) Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angewandte Chemie (International ed. in English)*, **51** (29), 7181–7184.

62 Winn M., Fyans J.K., Zhuo Y., Micklefield J. (2016) Recent advances in engineering nonribosomal peptide assembly lines. Natural Product Reports, **33**(2):317-347.

63 Broderick, J.B., Duffus, B.R., Duschene, K.S., Shepard, E.M. (2014) Radical S-adenosylmethionine enzymes. *Chemical reviews*, **114** (8), 4229–4317.

64 Leete, E., Davis, G.E., Hutchinson, C.R., Woo, K.W., Chedekel, M.R. (1974) Biosynthesis of azetidine-2-carboxylic acid in *Convallaria majalis*. *Phytochemistry*, **13** (2), 427–433.

65 Leete, E. (1975) Biosynthesis of azetidine-2-carboxylic acid from methionine in *Nicotiana tabacum*. *Phytochemistry*, **14** (9), 1983–1984.

66 Leete, E., Louters, L.L., Prakash Rao, H.S. (1986) Biosynthesis of azetidine-2-carboxylic acid in *Convallaria majalis*: Studies with N-15 labelled precursors. *Phytochemistry*, **25** (12), 2753–2758.

67 Ma, J.F., Shinada, T., Matsuda, C., Nomoto, K. (1995) Biosynthesis of Phytosiderophores, Mugineic Acids, Associated with Methionine Cycling. *Journal of Biological Chemistry*, **270** (28), 16549–16554.

68 Suzuki, M., Takahashi, M., Tsukamoto, T., Watanabe, S., Matsuhashi, S., Yazaki, J., Kishimoto, N., Kikuchi, S., Nakanishi, H., Mori, S., Nishizawa, N.K. (2006) Biosynthesis and secretion of mugineic acid family phytosiderophores in zinc-deficient barley. *The Plant Journal: for cell and molecular biology*, **48** (1), 85–97.