Ecole Doctorale SIReNa (Sciences et Ingénierie des Ressources Naturelles)

Thèse

Présentée et soutenue à huis clos pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention : « Ecotoxicologie, Biodiversité, Ecosystèmes »

Li SU

Generation of analogues of the anti-tumor polyketide stambomycins by genetic engineering and allied approaches

Synthèse d'analogues de stambomycines, une famille de polycétides présentant des activités antitumorales, par ingénierie génétique et approches associées

4 May 2021

Membres du jury :

Rapporteurs :

Monsieur Barrie WILKINSON	Professeur, John Innes Centre, Royaume Uni
Madame Cécile BON	Maître de conférences, HDR, Université Toulouse III − Paul Sabatier

Examinateurs :

Madame Yanyan LI	CR1, CNRS-Muséum National d'Histoire Naturelle, Paris
Monsieur Stéphane DESOBRY	Professeur, Université de Lorraine
Monsieur Russell COX	Professeur, Leibniz Universität Hannover
Monsieur Bertrand AIGLE	Professeur, Université de Lorraine (Directeur de thèse)
Madame Kira WEISSMAN	Professeur, Université de Lorraine (Co-directrice de thèse)
Monsieur Christophe JACOB	Maitre de conférences, HDR, Université de Lorraine

---

UMR UL-INRAE 1128 Dynamique des Génomes et Adaptation Microbienne (DynAMic)
Université de Lorraine, Faculté des Sciences et Technologies − Campus Aiguillettes − BP 239, 54506 Vandœuvre-lès-Nancy Cedex
UMR 7365 CNRS-UL, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA)
Université de Lorraine, Faculté de médecine, Bâtiment Biopôle, 9 avenue de la Forêt de Haye, BP 20199, 54505 Vandœuvre-lès-Nancy Cedex

Acknowledgements

First, I would like to thank my advisors Prof. Kira WEISSMAN and Prof. Bertrand AIGLE for their continual support of my research, my growth as a scientist. None of the work would be possible without their mentorship.

I would like to acknowledge all members of DynAMic (Team StrAda) and IMoPA (Team 3-Molecular and Structural Enzymology), past and present. Everyone in the two labs has positive effect on my research and success through helpful discussion. I would like to specifically thank Dr. Christophe JACOB for his help in guiding me the CRISPR-Cas9 mediated genetic engineering, and Laurence HÔTEL for her help in getting me started in working with *S. ambofaciens*. More importantly, Christophe and Laurence made some important constructs for this study before I involved. I would also like to thank Dr. Yaouba SOUAIBOU for his efforts and contributions to certain subprojects. A lot of thanks to Dr. Benjamin CHAGOT and Sabrina COLLIN for their kind help on *in vitro* study of 'DD'.

I would like to thank Prof. Jörn PIEL and Dr. Alexander O. BRACHMANN for welcoming me in his lab and for all the precious help in metabolites profiling of engineering strains. I enjoyed my short-staying in Zürich very much.

I would also like to acknowledge Monsieur Russell COX and Madame Yanyan LI for serving as members of my committee (comité de suivi individuel) in the past three years, and for their valuable opinions on the content and direction of my research.

Additionally, I would like to thank my previous advisor (Master research), Prof. Yi YU, for all his support in training and preparing me for a career in research.

Again, a special thanks to Kira, Bertrand and Christophe for reading and correcting my manuscript, you gave me very useful advice and comments. In addition, a lot of extra thanks to Bertrand for writing the special chapter "Summary in French".

Finally, I would like to thank Monsieur Barrie WILKINSON and Madame Cécile BON for accepting to be my referees, and Madame Yanyan LI, Monsieur Stéphane DESOBRY, Monsieur Russell COX to be the examiner. For me it has been a great honor and I hope for them it will be a pleasure to read this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

Chapter IV

Appendix to chapter II

Appendix to chapter III

Appendix to chapter IV

LIST OF TABLES

Chapter II

Chapter III

Appendix of Chapter II

Appendix of Chapter III

Appendix of Chapter IV

INDEX OF ACRONYMS AND ABBREVIATIONS

NP ..... natural products
SMBGC ..... secondary metabolites biosynthetic gene cluster
BGC ..... biosynthetic gene cluster
ORF ..... open reading frame
OSMAC ..... one stain-many compounds
PDB ..... precursor-directed biosynthesis
PKS ..... polyketide synthase
FAS ..... fatty acid synthase
CoA ..... coenzyme A
SNAC ..... N-acetylcysteamine
LM ..... loading module
ACP ..... acyl carrier protein
AT ..... acyltransferase
DH ..... dehydratase
ER ..... enoyl reductase
KS ..... ketosynthase
KR ..... ketoreductase
TE ..... thioesterase
DD ..... docking domain
KAL ..... KS-AT interdomain linker
PAL ..... post-AT linker
DEBS ..... 6-deoxyerythronolide B synthase
6-dEB ..... 6-deoxyerythronolide B
GT ..... glycosyltransferase
MT ..... methyltransferase
MAT ..... malonyl-CoA:ACP transacylase
MatB ..... malonyl-CoA synthetase
ACC ..... acetyl-CoA carboxylase
PCC ..... propionyl-CoA carboxylase
CCR ..... crotonyl-CoA carboxylase/reductase
ATP ..... adenosine triphosphate
PPTase ..... phosphopantetheinyltransferase
RAPS ..... rapamycin synthase
Pik ..... pikromycin synthase
Ole ..... oleandolide synthase
LAL ..... Large ATP-binding LuxR
GNPS ..... Global Natural Products Social Molecular Networking-platform
DFO ..... desferrioxamine
FO ..... ferrioxamine
BAC ..... bacterial artificial chromosome
CRISPR ..... clustered regularly interspaced short palindromic repeats
RBS ..... ribosome binding site
FRT ..... FLP recognition targets
HAL ..... homologous arm left
HAR ..... homologous arm right
rpm ..... revolutions per minute
FPLC ..... fast protein liquid chromatography
HPLC ..... high-performance liquid chromatography

HR-LCMS ..... high-resolution liquid chromatography mass spectrometry
IPTG ..... isopropyl-β-D-thiogalactopyranoside
MWCO ..... molecular weight cut-off
$NADP^+$ ..... nicotinamide adenine dinucleotide phosphate, oxidized
NADPH ..... nicotinamide adenine dinucleotide phosphate, reduced
H3C protease ..... human rhinovirus 3C protease
SDS-PAGE ..... dodecyl sulfate polyacrylamide gel electrophoresis
ITC ..... isothermal titration calorimetry
CD ..... circular dichroism
SEC-MALS ..... size-exclusion chromatography multiangle light scattering
TCEP ..... tris(2-carboxyethyl) phosphine
3-HBA 3-hydroxybenzonate
3,5-DHBA 3,5-dihydroxybenzonate
DHCHC ..... 5-dihydroxycyclohex-1-enecarboxylic acid
SNAC ..... synthetic N-acetylcysteamine
SDM ..... site directed mutagenesis
NRPS ..... non-ribosomal peptide synthetases
NIS ..... NRPS-independent siderophore
FDA ..... Food and Drug Administration
HSC ..... N-hydroxy-N-succinylcadaverine
HAC ..... N-hydroxy-N-acetylcadaverine
LDC ..... L-lysine decarboxylation
DABA DC ..... L-2,4-diaminobutyrate decarboxylase
DABA AT ..... L-2,4-diaminobutyrate acyltransferase

Chapter I Introduction

1.1 General context

Until recent times, infective diseases and cancer were two leading causes of death in humans. Unfortunately, loss of drug effectiveness due to antimicrobial resistance (AMR) [1] and cancer cell resistance against chemotherapeutic agents [2] is increasing in both developing and developed countries. In addition to the impact on human health, infection disease and cancer are two of the critical issues causing economic and financial burden in the world today. For example, the European Centre for Disease Prevention and Control (ECDC) has estimated that the cost of infections caused by antimicrobial resistant bacteria in the EU amount to almost 1.5 billion per year over the last decade [1]. Similarly, the world cancer report demonstrated that the total annual economic cost of cancer is estimated at approximately US$1.16 trillion [3]. In this formidable context, the challenge is to discover novel and effective anti-infective agents possessing useful pharmaceutical properties, including high target specificity combined with a good bioavailability, and low or no toxicity.

Polyketides are secondary bioactive metabolites synthesized by enzyme complex (polyketide synthase, PKS). Because of the complex biosynthetic machinery, these molecules have complex and diverse structures, and are generally reported having antibacterial (e.g. erythromycin [4], spiramycin [5]) or anticancer bioactivities (e.g. epothilone B [6], marmycins [7]), warranting therapeutic potential for antibacterial and anticancer treatments. Although the native polyketides can in some cases be utilized directly as drug (e.g. erythromycin A (https://clincalc.com/DrugStats/Drugs/Erythromycin), structural alteration of the molecules is typically necessary in order to tailor their therapeutic (ADMET) properties.

This PhD project endeavored to develop synthetic biology approaches towards generating novel compounds, using the polyketide stambomycins of *Streptomyces ambofaciens* ATCC23877 as a model system. Although the metabolites exhibit promising anti-cancer activities and are also weakly anti-bacterial, their use as drugs is limited by the size of the molecules (at 51-members, they are among the largest of all known polyketides),

and by their toxicity. In this context, the specific aim was to develop methods for accessing smaller derivatives which would potentially retain bioactivity, as well as structures modified at a key ring position directly impacting anti-proliferative potency, and to evaluate the resulting structures for their toxicity/therapeutic properties.

1.2 Natural products, microbial secondary metabolites, and antibiotics

Natural products (NPs) usually exhibits some kinds of biological activities, including antibacterial and antifungal (e.g., erythromycin, candicidin, etc.), antitumor drugs (e.g., doxorubicin, daunorubicin, etc.), immunosuppressants (rapamycin, FK506, cyclosporin, etc.), enzyme inhibitors (clavulanic acid, serpins, etc.), antiparasitic agents (avermectins, milbemycins, cyclic peptides, etc.), bioherbicides (glufosinate, etc.) [8]. Of the numerous bioactive compounds, those acting as antimicrobials (antibacterial, antifungal and antiprotozoal) are termed antibiotics [9].

The interest for antibiotic research since the discovery of penicillin by Alexander Fleming in 1928 and its development in the early 1940s at Oxford by Chain and Florey [10], has been particularly strong, especially during the period 1940−1980, also known as the Golden Age of antibiotic discovery [11]. During this time, many antibiotics were discovered and used clinically, including the antibacterial agents tetracycline (1948), erythromycin (1952) and rifamycin (1965), and the antifungal agent candicidin (1953). However, the process of discovering these bioactive NPs is generally long and laborious, and often leads to the rediscovery of already known compounds. NPs also compete with the products of chemical synthesis massive chemical libraries can be synthesized and screened for a particular therapeutic activity. Indeed, this idea resulted in a wave of combinatorial chemistry initiatives in industry in the mid 1980's, as a means to access chemical diversity. As a result, the rate of antibiotic discovery slowed down gradually from 1980.

However, by the early 2000s, it was clear to many companies that despite a major investment in combinatorial chemistry, the approach failed to deliver on its promises. Indeed, synthesizing a large compound library does not in itself increase the odds of discovering a preclinical candidate, as drug discovery is not a simple number games [12]. As a result, interest returned to the privileged NP structures, which relative to the majority

of synthetic compounds, exhibit higher chemical diversity, structural complexity and biological potency [doi: 10.1038/nature03194]. Simultaneously, with the development of DNA sequencing technology, a large number of genome sequences of microbes have been published and can be freely obtained, facilitating 'mining' of specialized secondary metabolite biosynthetic gene clusters (SMBGCs). Indeed, interest in discovering new NPs based on genome mining is on the rise and has received much attention over the last several years [13-19].

1.3 *Streptomyces*, an attractive organism for discovery of natural products

Streptomyces, one genus of the family *Streptomycetaceae*, contains multiple species that are abundant in soil. *Streptomyces* are unique among soil bacteria in that they grow as branching hyphal filaments to form a mat of fungus-like mycelium, from which emerge aerial branches that bear chains of spores [20]. When a spore finds favorable conditions in terms of temperature, nutrients, and moisture, the germ tube is formed, and then the new cycle starts again (**Figure 1.1**). This growth adaptation assists in adhering to and penetrating the insoluble organic remains of fungi, plants, and other organisms, which are then broken down by hydrolytic exoenzymes (such as chitinases, amylases, proteases and cellulases), to provide nutrients, suggesting the key role of *Streptomyces* in the solubilization of cell wall and surface components of plants, fungi, and insects [21]. However, this is only one aspect of the interest and importance of streptomycetes.

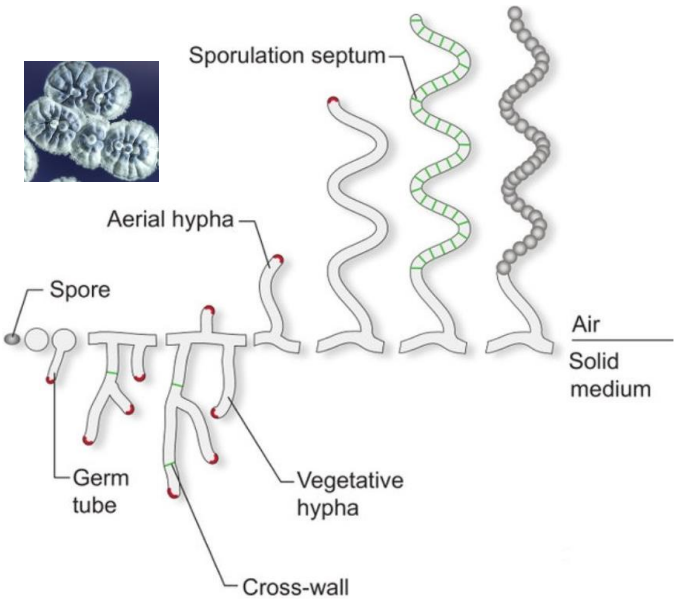


Figure 1.1 The *Streptomyces* life-cycle on solid media. The cellular development of a spore begins with the formation of one or two germ tubes, which grow by tip extension, separating into compartments connected by cross-walls, to form a network of branching hyphae. In response to nutrient limitation and other signals, aerial hyphae are erected. The aerial hyphae grow, forming fluff-like colonies, and after hyphae septation and spore maturation, spores are released to restart the life-cycle. Adapted from reference [22].

It has been estimated that over half of the antibiotics are produced by actinobacteria, 10–15% by non-filamentous bacteria (e.g., *Bacillus* and *Pseudomonas*) and about 20% by filamentous fungi. Among the species of actinobacteria, streptomycetes are the main producers, being responsible for 76% of all bioactive metabolites derived from actinobacteria [23].

Genome sequencing of *Streptomyces coelicolor* A3(2), a model streptomycete, revealed an 8,667,507 base pair linear chromosome [24]. The genome comprises 7825 predicted open reading frames (ORFs), encompassing more than 20 SMBGCs encoding for known or predicted specialized metabolites. As the genomes of increasing numbers of *Streptomyces* species were sequenced, researchers found that dozens of SMBGCs exist in almost every species, showing the promise of streptomycetes as an important microbial source for the production of natural products. In addition, the genome of *Streptomyces coelicolor* A3(2) encodes an extraordinary 819 predicted secreted proteins, including proteases, chitinases/chitosanases, pectate lyases etc., which comprises a comprehensive extracellular system, reflecting the close association of *Streptomyces* with other organisms.

Furthermore, the genome also contains 965 regulatory genes, accounting for 12.3% of the edsartntire genome, which presumably promote its survival in the highly competitive soil environment [24]. *S. coelicolor* has been extensively studied to understand the physiology, genetics, regulation and specialized metabolism of this genus, and also been developed as a heterologous host for a number of SMBGCs [25].

1.4 Strategies for novel natural products discovery

Traditionally, natural products were discovered through isolation of microorganisms, followed by large-scale fermentation, extraction, screening for biological activities, and separation of active components. These methods have led to the identification of many active NPs, including antibiotics (e.g., streptomycin). However, this approach is limited by the fact that some producing strains exhibit unfavorable characteristics, such as slow growth rate and low production yields, particularly under classic laboratory conditions. In addition, these efforts were hampered by rediscovery of the same metabolites, which led to the erroneous conclusion that Nature's compound reservoir had been fully tapped. With the advent of wide-scale genome sequencing, as mentioned earlier, we now know that the microbial genetic potential for specialized metabolite synthesis far exceeds what is observed in the laboratory. In addition, it's been estimated that only 1% of microorganisms are culturable under laboratory conditions, meaning that more than 99% of microbes have not been explored [26]. In order to make better use of these natural resources for the benefit of mankind, scientists urgently need to continue to develop effective approaches for discovering new NPs. Over the past two decades, various strategies have been established and applied to discover natural specialized metabolites, as well as to create new "unnatural natural" products.

1.4.1 OSMAC and co-cultivation

The production of specialized metabolites by microorganisms is strongly dependent on environmental factors, such as growth conditions and biotic stresses. This observation inspired the one strain many compounds (OSMAC) approach [27], in which culture parameters (media, nutrients, oxygenation, etc.), are systematically varied to simulate different environmental conditions. This approach has led in a number of cases to the identification of novel specialized metabolites [28]. Another related approach to stimulate

the biosynthesis of secondary metabolites is to challenge a given microorganism with biotic stress by co-cultivation. In nature, microorganisms are often found in communities, where they produce specialized metabolites which function as defensive compounds or as signaling molecules in ecological interactions, nutrients competition, metal transportation, etc. [29]. Such situations can be reproduced artificially, to some extent, by culturing two or more microorganisms together. Application of the OSMAC and co-cultivation strategies has resulted in the discovery of many novel natural products, such as new manumycins and manumycin analogues when the manumycin producer *Streptomyces parvulus* Tü64 was fermented under increased pressure [30], and the antibiotic fumicycline when *Aspergillus fumigatus* was co-cultivated with *Streptomyces rapamycinicus* [31]. More comprehensive descriptions and examples are provided in reviews by Bertrand *et al*., De Roy *et al.*, and Marmann *et al*. [32-34].

1.4.2 Genome-directed strategies

In the early 2000, the genome sequencing of *S. coelicolor* A3(2) was completed and it became evident that streptomycetes carry the genetical potential to produce many more natural products than those detected under laboratory conditions [24]. Since that time, interest in discovering novel natural products and generating unnatural natural products by genome-directed approaches has been growing apace, via strategies including genome mining, pathway engineering, combinatorial biosynthesis, precursor-directed biosynthesis and mutasynthesis. An overview of these strategies towards the generation of novel natural products is shown in **Figure 1.1**.

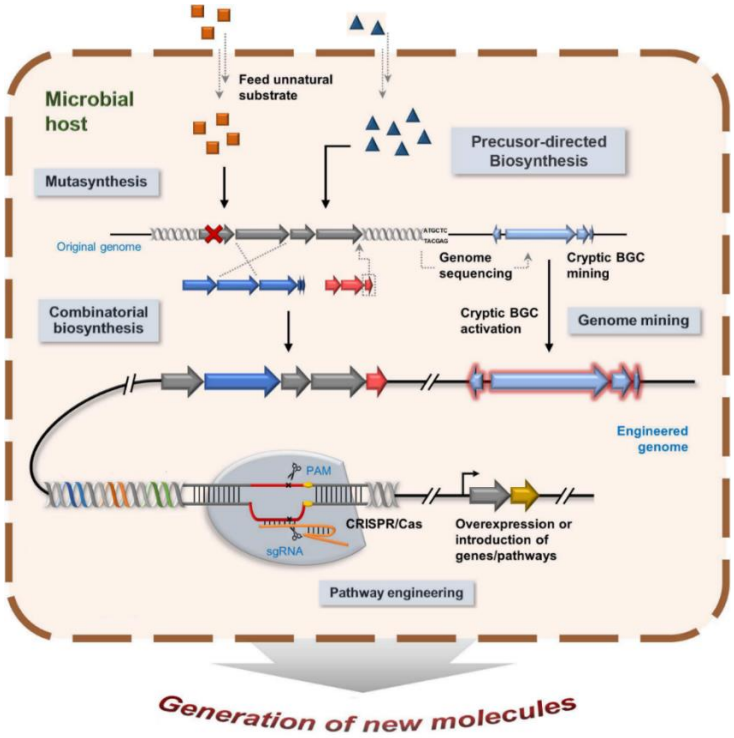


Figure 1.2 An overview of multiple genome-directed strategies towards the generating new natural products. Adapted from reference [35].

1.4.2.1 Genome mining

Genome mining is an approach to discover novel specialized metabolites which involves extracting information from genome sequencing. As described earlier, each streptomycete has the ability to carry more than 20 SMBGCs, many of their encoding metabolites remain unidentified. With the wealth of genetic data and a suite of bioinformatic tools available (e.g. antiSMASH [36], MIBiG platform [37], etc.), researchers are now able to computationally mine the genetic data, and either connect them to known secondary metabolites or predict the structures of novel compounds [38, 39]. An elegant example of this strategy is the identification and characterization of the stambomycin family of compounds. In this case, Laureti and co-workers found a cryptic gene cluster encoding the biosynthesis of an unknown polyketide in the genome of *Streptomyces ambofaciens* ATCC23877. Following analysis of the transcriptional regulatory network, they identified the BGC product by constitutive expression of a cluster-specific activator SamR0484, belonging to the LAL (Large ATP-binding regulator of the LuxR) family [40] (for more details, see **section 1.8**). A similar example is the discovery and characterization of

neoansamycin from *Streptomyces* sp. LZ35 [41]. Initially, genome sequence analysis revealed the presence of a putative ansamycin gene cluster (*nam*), but reverse transcription (RT)-PCR analysis showed that the *nam* gene cluster was only weakly expressed under conventional laboratory conditions. To induce the expression of the *nam* gene cluster, the constitutive *ermE** promoter was inserted upstream of the gene *nam1*, which encodes a putative positive regulator of the LuxR family proteins. The constitutive expression of gene *nam1* successfully activated the *nam* gene cluster, and led to the isolation of three novel ansamycins, namely neoansamycins A–C (**Figure 1.2**).

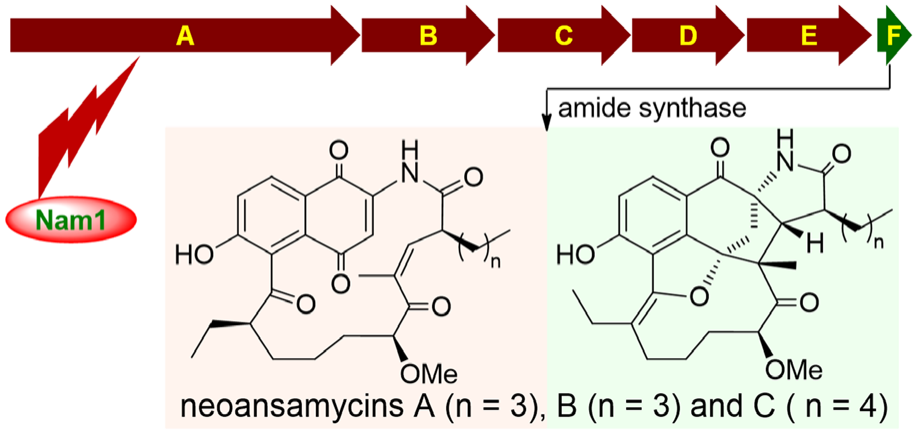


Figure 1.3 Activating a cryptic ansamycin biosynthetic gene cluster to produce three new ansamycins with *n*-pentyl and *n*-butyl side chains. Adapted from reference [41].

1.4.2.2 Pathway Engineering and Combinatorial Biosynthesis

In microorganisms, the most common pattern of chemical diversification is the horizontal transfer entire gene clusters or portions thereof, which then evolve differently by recruitment of new genes and loss of others. Consistent with the proposed model of evolution are the many families of natural products with common structural cores, but bearing different tailoring modifications [42]. Commonly encountered modifying enzymes carry out reactions such as glycosylation, acylation, oxidation, methylation, etc. In this context, one engineering strategy is to target the late stage modifications, while leaving the central scaffold intact, although this invariably leads to structurally-related compounds. For example, the aminocoumarins novobiocin and clorobiocin share a common structural core but differ at the C-8 position of the aminocoumarin moiety by a methyl or a chlorine

group, respectively, and at the 3-OH position of the sugar moiety, by a carbamoyl or a methyl-pyrrole-2-carboxyl group. By replacing the methyltransferase in the novobiocin pathway with the halogenase from the clorobiocin pathway and *vice versa*, a series of new hybrid aminocoumarins was produced (**Figure 1.3**) [42, 43]. This pathway engineering approach focused on tailoring enzymes has also been widely applied to structurally diversify polyketides (for details, see **section 1.7.5**).

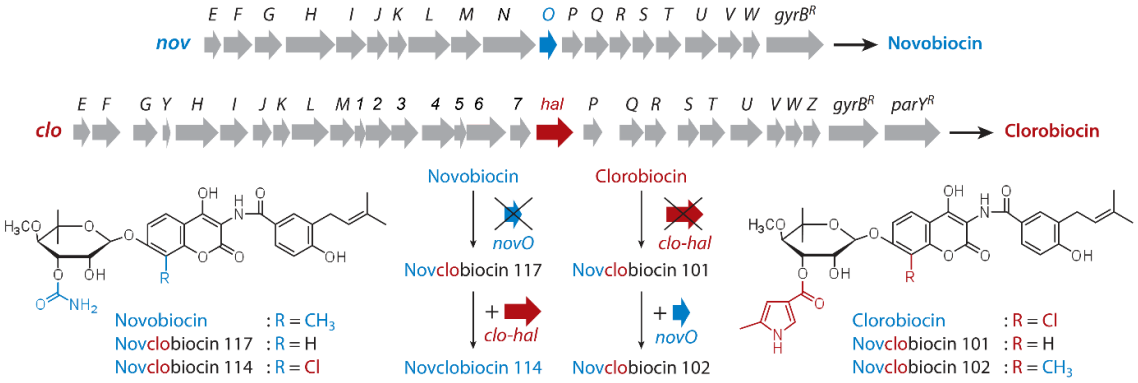


Figure 1.4 Pathway engineering to produce novobiocin/clorobiocin analogues by exchange of tailoring genes. The *nov* and *clo* gene clusters are shown in parallel to illustrate the homology and divergence of the two pathways. *clo-hal* encodes the halogenase and *novO* encodes the methyltransferase. Adapted from reference [42].

Alternatively, or in parallel, it is possible to modify the core structures. Indeed, certain classes of natural product (polyketides and nonribosomal peptides) are at least in principle particularly amenable to this approach, as they are biosynthesized in assembly line fashion by gigantic multienzymes. Thus, changes introduced at the genetic level can yield predictable outcomes in terms of product structure. Theoretically, generation of analogues can be achieved by mixing and matching of megasynthases at the subunit, module, and domain levels. This point will be returned to in **section 1.7**.

1.4.2.3 Precursor-directed biosynthesis and mutasynthesis

Precursor-directed biosynthesis (PDB, **Figure 1.4a**) is performed by the addition of analogues of biosynthetic precursor into the culture medium. If the organism is then capable of incorporating the precursor into the native pathway, a modified metabolite is obtained [44]. Mutasynthesis (*muta*tional bio*synthesis*, **Figure 1.4b**) extends the PDB approach by inactivating the pathways leading to competing precursors [44]. The technique

is based on the inactivation of one or more of the early-stage enzyme activities, such that biosynthesis of the target NP by the resulting mutant is, at least in theory (see **section 3.3.2**), blocked. Production can then be restored by provision of analogues of native building blocks or a more advanced intermediates, again giving rise to derivatives of the natural product.

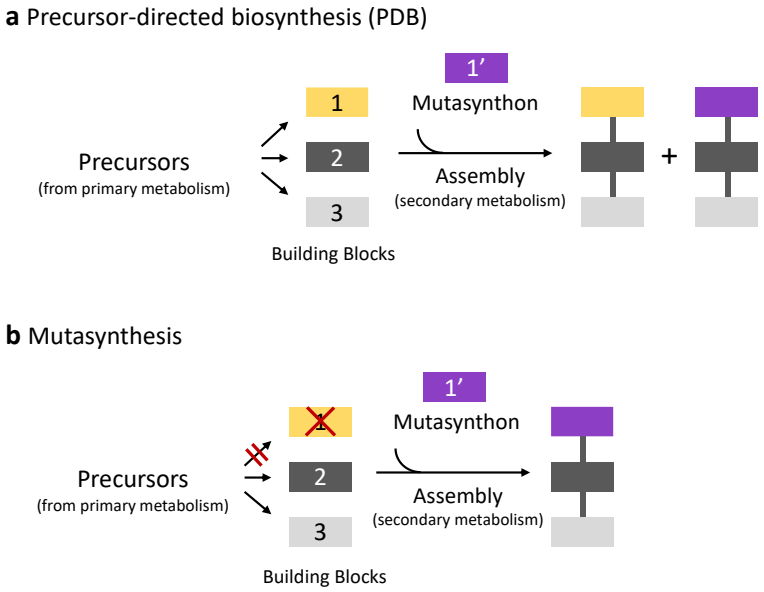


Figure 1.5 Schematic illustration of precursor-directed biosynthesis (PDB) and mutasynthesis. a) In precursor-directed biosynthesis, the culture medium is supplemented with an analogue of the natural building block, which competes for incorporation into the natural products. Both native and modified products are usually produced. **b**) In mutasynthesis, biosynthesis by the non-producing mutant strain is reactivated by addition of the feeding molecules and only novel analogues are produced. Adapted from reference [44].

The three main requirements for successful PDB and mutasynthesis are: i) the analogue to be introduced must be chemically stable under the conditions used; ii) the analogue must be recognized by the biosynthetic machinery, and incorporated into the structure; and iii) the resulting modified intermediate(s) must be accepted and processed by the remaining downstream enzymes in the pathway. It is worth mentioning that PDB and mutasynthesis have already been applied with success to diversify polyketide structures, a theme which will be further developed in **section 1.7.6**.

1.5 Polyketide natural products

Polyketides are a large class of natural products from producing organisms as diverse as bacteria, fungi, plants, insects, dinoflagellates, sponges, and animals, and which exhibit a vast array of biological and pharmacological activities, including antibacterial, antifungal, anticancer and immunosuppressive properties [45-47]. As a testimony to their importance, the market value for polyketide-derived medicines routinely reaches 20 billion US dollars per year [48]. Examples of clinically useful polyketides are presented in **Figure 1.5**.

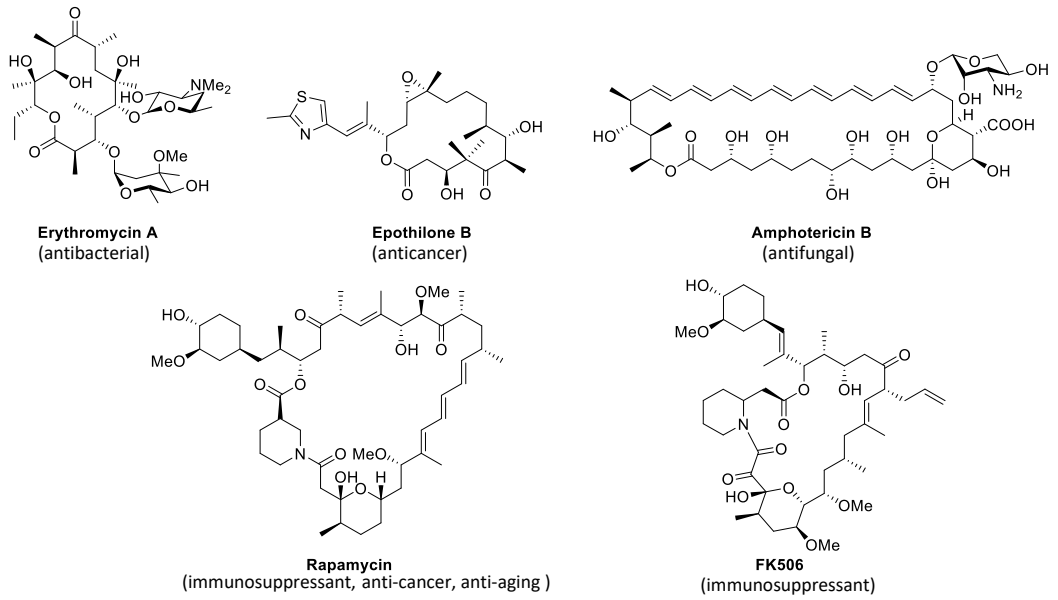


Figure 1.6 Examples of therapeutically relevant polyketide natural products.

1.6 Polyketide synthases (PKSs)

1.6.1 Classification of PKSs

On the basis of sequence, architecture, and catalytic mechanism, PKSs are now classified into three groups: type I, type II and type III.

1.6.1.1 Modular type I polyketide synthases and iterative type I polyketide synthases

Type I PKS refers to linearly arranged and covalently fused catalytic domains within large multifunctional enzymes, which can be further subdivided into modular and iterative types. As the name indicates, modular type I PKSs are composed of multiple modules where each module consists of all the necessary domains required for a single round of polyketide chain elongation and associated carbon processing. Type I iterative PKSs, on the other

hand, utilize a single set of covalently bound catalytic domains iteratively until the required length of the polyketide product with correct carbon processing is reached [49].

Type I modular PKSs can also be further classified as *cis*-acyl transferase (AT) and *trans*-AT (or AT-less) systems. *cis*-AT systems (e.g. deoxyerythronolide B synthase (DEBS), **Figure 1.7a**) incorporate an AT domain in *cis* within each module, which is responsible for loading the specific starter unit or extender unit onto its cognate module. In *trans*-AT systems (e.g. the anthracimycin, virginiamycin, lankacidin, kirromycin PKSs, etc.), in contrast, the AT domain is not located within each module, but exists as a discrete enzyme which acts iteratively to load extender units onto the chain extension modules (**Figure 1.7b**).

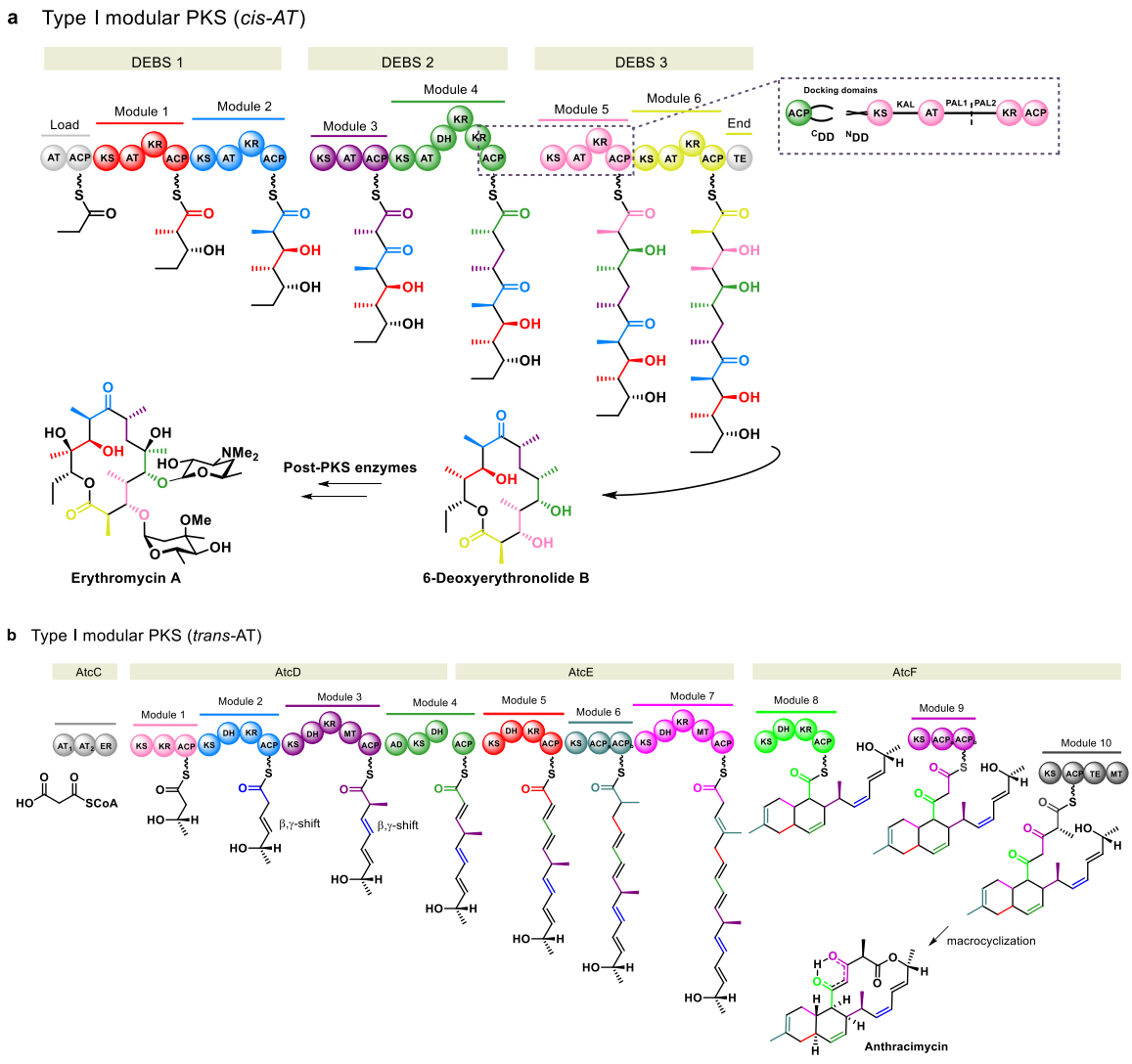
The DEBS system of *Saccharopolyspora erythraea* was the first type I PKS to be described and continues to serve as a model for understanding the structure and biochemical function of modular PKSs. This PKS system consists of three gigantic polypeptides (called 'subunits'), each of which is composed of modules – the first which initiates the biosynthesis (load), and six to extend the chain. Each of the chain elongation modules comprises three core domains: AT (acyltransferase), KS (ketosynthase) and ACP (acyl carrier protein) [50]. The AT domains load the starter unit (e.g. acetyl-CoA) and extender units (e.g. malonyl-CoA) onto their partner ACP domains, via covalent tethering to the terminal thiol of a phosphopantetheine (Ppant) prosthetic group. These starter unit/chain extension intermediates are elongated by the KS domains, which catalyze Claisen-like condensation reactions using a nucleophilic enolate derived from decarboxylation of the extender unit.

Apart from these three core domains, the modules additionally contain other types of catalytic activities, which stereospecifically modify the oxidation state of the initially formed β-ketothioester. When present, a ketoreductase (KR) domain reduces the α-C=O to the α-CH-OH, a dehydratase (DH) domain eliminates H_2O to generate an α, β-double bond (α-C=C-β), and an enoyl reductase (ER) domain reduces the α, β-double bond to a saturated methylene (α-CH_2-CH_2-β) [51, 52]. The modules themselves contain multiple catalytic and carrier protein domains, separated by so-called 'linker' regions. These linkers are typically 20 residues in length, but may include as many as 250 residues [53] (**Figure**

1.7a insertion box, KAL, PAL1 and PAL2). They play an important architectural role in the biosynthetic process [49], by acting as rigid interdomain adaptor regions and/or by conferring flexibility to the synthases, thereby facilitating the movement of the domains within the multienzymes [54]. In addition, docking domains located at the extremity of the PKS proteins (**Figure 1.7a** insertion box), which are present in the form of folded structures, act to ensure that interactions between the successive subunits occur correctly [55] (see more details in **section 1.7.3**). At the end of the synthesis, the full-length acyl chain is released from the assembly line by macrolactonization by a thioesterase (TE) domain (although in other systems, alternative hydrolytic liberation occurs [56, 57]). The compound released from the PKS, 6-deoxyerythronolide B, is not the final product of the pathways, as it is then set upon by a set of discrete 'post-PKS' enzymes including two P450 hydroxylases, two glycosyltransferases (GTs) and a methyltransferase (MT) to afford erythromycin A. Such enzymes are typical of the post-PKS processing phase of the biosynthesis, which is often necessary to afford the final biologically-active forms of the metabolites.

Despite the presumed distinct evolutionary origin of the *trans*-AT PKS systems [58], the biosynthesis occurs by a similar overall logic. Anthracimycin, for example, is a 14-membered macrolide antibiotic assembled by a *trans*-AT type I PKS system in the marine strain *Streptomyces sp. T676.* Analysis of the *atc* cluster revealed that anthracimycin is generated by four PKS gene products AtcC−AtcF (**Figure 1.7b**) [59]. AtcC contains two AT domains and one ER domain, in which the AT_1 acts as the acyl-transferase *in trans* to load building blocks (malonyl-CoA) and the AT_2 acts as an acyl hydrolase for the removal of aberrant or stalled acyl units from blocked *trans*-AT PKS modules. The C-terminal ER acts *in trans* to reduce the enoyl bond in certain steps. The following ActD–F contains ten modules, which are responsible for the elongation of the polyketide chain. The biosynthetic route was proposed to be initiated by loading a malonyl-CoA by AT_1 followed by the decarboxylation and Claisen condensation by the KS domain of module 1. When the polyketide chain is extended by module 8, the resulting intermediate undergoes [4 + 2] cycloaddition, resulting in the decalin moiety. Following two further rounds of chain elongation, the TE domain catalyzes chain release with associated macrolactone formation, resulting in the final product anthracimycin. Notably, module 4 is split between two

proteins (AtcD and AtcE), a type of dissociated module organization which is common in *trans-AT* PKSs (e.g., in the difficidin, myxovirescin and lankacidin PKSs [60-62]). The author predicted that the split module 4 could be important in allowing the pentaketide intermediate to access either the KR_5 domain or the AD domain, which sits between modules 3 and 4 [59].



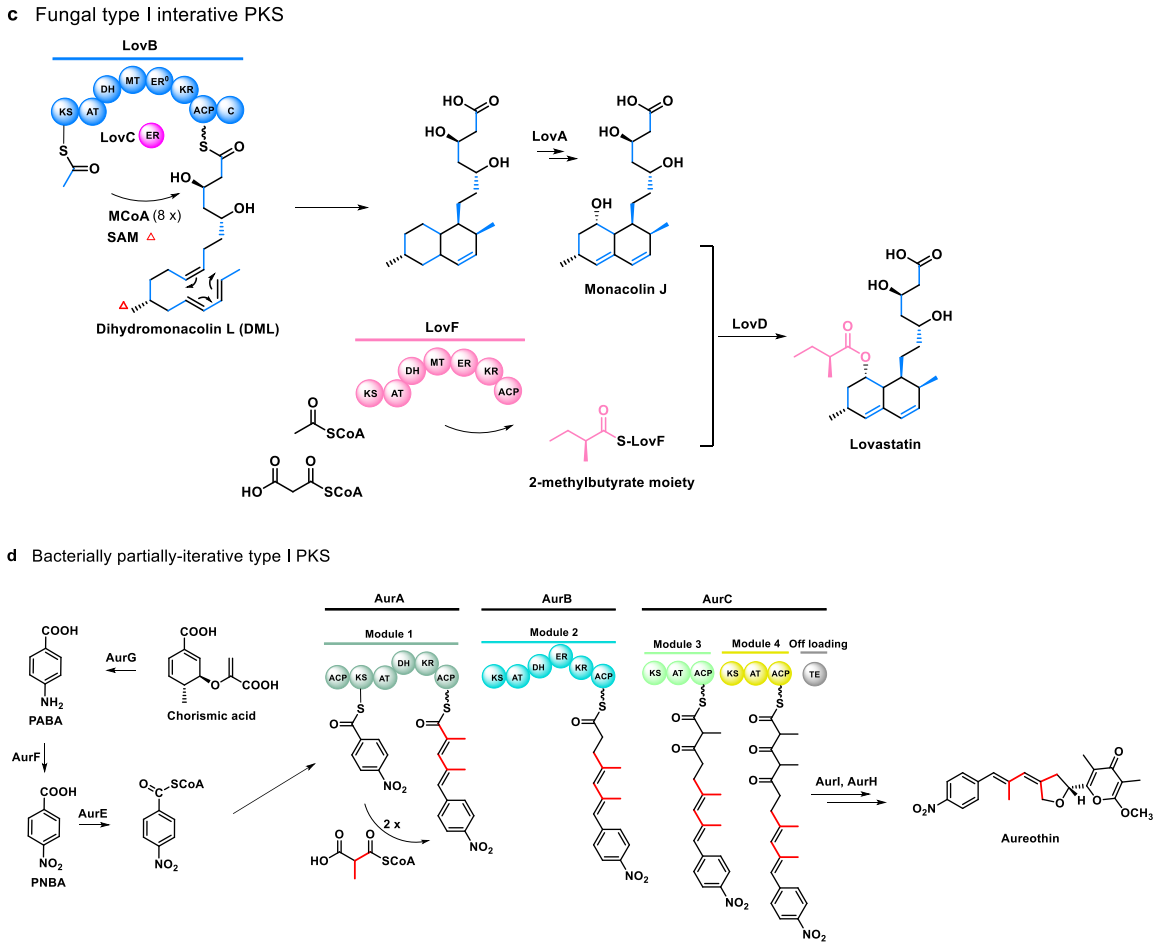


Figure 1.7 Type I PKSs exemplified by the erythromycin (*cis*-AT modular), anthracimycin (*trans*-AT modular), lovastatin and aureothin (mixed iterative/modular *cis*-AT PKS) system. Key to domains: AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase; MT, *C*-methyltransferase; ACP, acyl carrier protein; AD, alcohol dehydrogenase; ^{C}DD, C-terminal docking domain; ^{N}DD, N-terminal docking domain; KAL, KS to AT linker; PAL1 and PAL2, post-AT linkers. Adapted and revised based on references [59, 63-66].

The type I iterative PKSs are mainly found in fungi, as exemplified by the lovastatin system in *Aspergillus terreus* (**Figure 1.7c**). The lovastatin biosynthetic pathway consists of two PKSs LovB and LovF. LovB consists of a ketosynthase (KS), a malonyl transferase (AT), a methyltransferase (MT), a dehydratase (DH), an enoyl reductase (ER), a ketoreductase (KR), an acyl carrier protein (ACP) and a condensation domain (C). The ER domain in LovB is nonfunctional, but a free-standing ER LovC is utilized as a partner. LovB recruits 9 malonyl-CoA extender units to produce the nonaketide moiety dihydromonacolin L (DML). LovF consists of all the above-mentioned domains, but it has

a functional ER. LovF catalyzes the formation of the 2-methylbutyrate moiety by the condensation of two acetyl units. Other enzymes encoded within the same gene cluster include LovA and LovD. LovA is a cytochrome P450 oxygenase which oxidizes the DML and result in monacolin J, which is then covalently joined to the 2-methylbutyrate moiety by the acyltransferase LovD, giving rise to the final product lovastatin.

While type I iterative PKSs are characteristic of fungi, they are also present in bacteria. Bacterial iterative type I PKSs can be divided in two categories: entirely iterative PKSs (fungal iterative PKS-like, exemplified by AviM, a single-module PKS responsible for the biosynthesis of orsellinic acid, a phenolic moiety in avilamycin [67]) and partially iterative PKSs (module iteration, exemplified by the aureothin polyketide synthase, **Figure 1.7d** [68]). In the multi-modular PKS of aureothin, module AurA (or subunit, as only one module is present in the first subunit), works iteratively two catalyze two chain elongations. Hertweck and co-workers proposed that the iteration occurs through the retrotransfer of the biosynthetic intermediate from one subunit to the opposite subunit of the homodimeric AurA [69]. Following this iterative process, the chain is extended in simple, modular fashion with malonate (module 2) and two units of methylmalonate (modules 3 and 4). Finally, the assembled polyketide chain is subjected to further enzymatic tailoring by the *O*-methyltransferases (AurI) and the cytochrome P450 mono-oxygenase (AurH), resulting in the final product aureothin.

1.6.1.2 Type II PKSs

In contrast to the type I PKSs in which the catalytic domains are covalently linked into large multienzymes, the components of type II PKSs are discrete proteins which assemble into non-covalent complexes [64]. The catalytic centers act in a repetitive manner during successive cycles of polyketide chain extension. Type II PKSs produce aromatic, polycyclic polyketides, including the well-studied actinorhodin, tetracycline, doxorubicin, and frenolicin [70].

The 'minimal' iterative PKS (**Figure 1.8**) consists of a ketosynthase heterodimer (KS_α/KS_β), an acyl carrier protein (ACP) and a malonyl-CoA:ACP transacylase (MAT, which is recruited from bacterial fatty acid synthesis). By analogy to type II fatty acid biosynthesis, the key step in the formation of poly-β-keto chain is the decarboxylative

Claisen condensation of malonate extender units with an acyl starter. The KS_{α} of the 'minimal' type II PKS contains an active site cysteine and catalyzes the successive Claisen condensations which result in a poly-β-ketone chain, whereas the KS_{β}, for which the active site Cys is substituted by a glutamine residue, carries out decarboxylation of malonyl-ACP to generate the acetate starter unit, which is then transferred to KS_{α} to initiate the biosynthetic sequence. The KS_{β} enzymology is reminiscent of the KS^{Q}s found in the modular type I PKS, which similarly convert malonyl-ACP into acetyl-ACP within loading modules [71]. The KS_{β} is alternatively known as the chain length factor (CLF), as it has been demonstrated to play a major role in controlling the final length of the polyketide chain [72]. However, KS_{β} is not the sole determinant, as the entire PKS complex, and in particular cyclases (CYCs), appear to be involved in controlling the number of chain elongation cycles [73]. In some cases, specific carbonyl functions can be reduced by a ketoreductase (KR) with the regiochemistry of reduction influencing first ring formation. In general, the poly-β-ketone products of the type II PKSs can undergo intramolecular Claisen and aldol reactions of varied regiochemistry, and thus this mode of biosynthesis gives rise to a diversity of aromatic compounds [49].

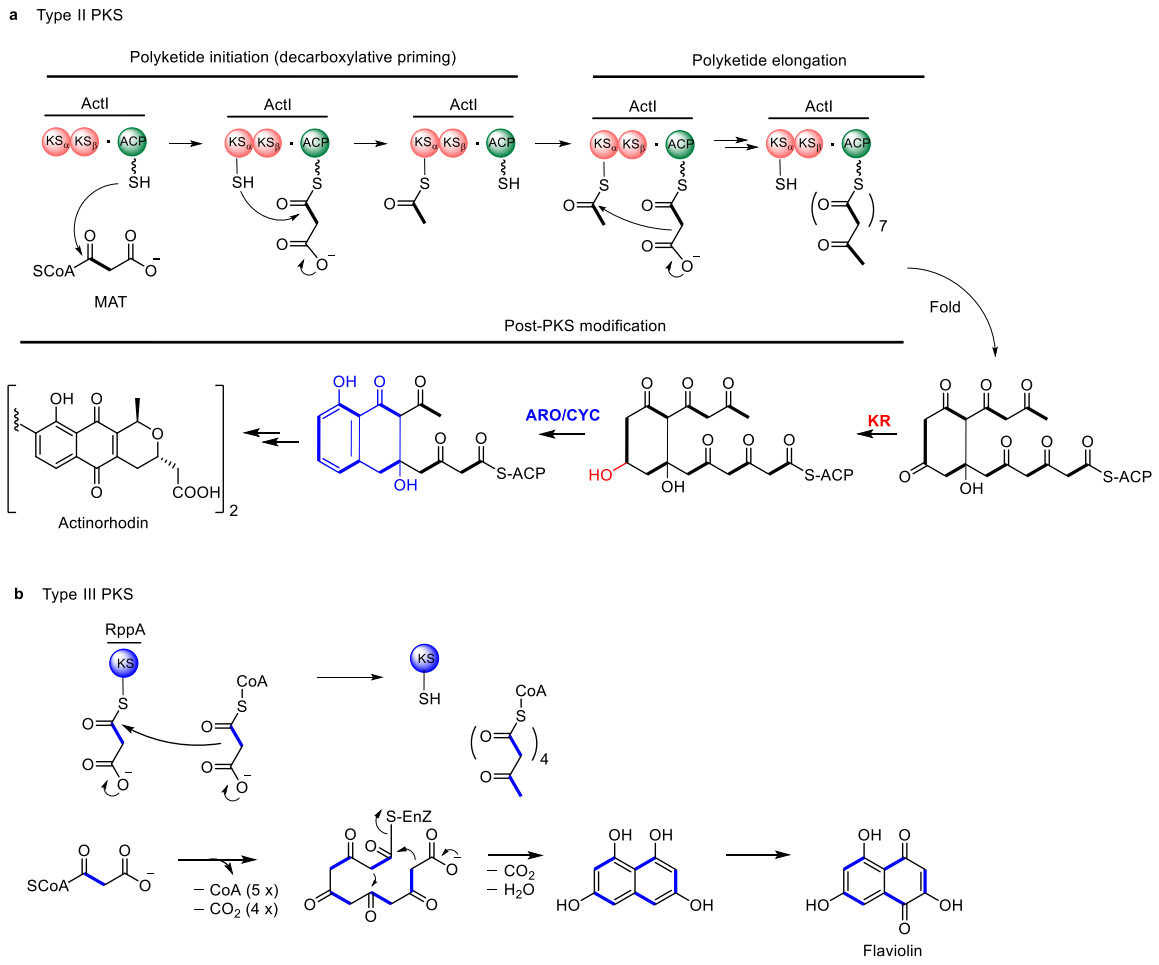


Figure 1.8 Examples of type II PKS and type III PKS involved in the biosynthesis of actinorhodin and flaviolin, respectively. a) The chain length factor (CLF also named KS_β as shown here) and ketosynthase (KS_α) form a heterodimer. A malonyl-CoA:ACP transacylase (MAT) primes the ACP with a malonate unit. The biosynthesis begins with the decarboxylation of a malonyl unit tethered to the KS_α, followed by Claisen condensation with an activated malonyl unit attached to the ACP. Seven rounds of Claisen condensation takes place overall to reach the required chain length. Upon release, the product undergoes processing by KR and CYC/ARO and additional enzymes to produce actinorhodin. **b**) Synthesis of flaviolin by a type III polyketide synthases/chalcone synthase. The growing chain is shown attached to a single KS-like enzyme (RppA). The malonyl-CoA starter unit is condensed iteratively with four additional malonyl-CoA units, followed by aromatization to the final product. Adapted from reference [64].

1.6.1.3 Type III PKSs

This type of polyketide synthase is commonly found in plants where it gives rise to chalcones and stilbenes, but has more recently been discovered in bacteria (e.g. flaviolin and phloroglucinol [74, 75]). Chalcone synthase was the first type III enzyme discovered

in the biosynthetic pathway to the flavonoids, followed shortly thereafter by the stilbene synthase, leading to these type III PKSs being collectively termed the 'chalcone synthase/stilbene synthase superfamily' (CHS/STS) [76]. Like type I and type II PKSs, type III PKSs condense a starter unit with a series of extender units to generate a polyketide chain. However, there are multiple differences between these classes, including the lack in type III PKSs of multiple catalytic domains for chain building, and the use of free CoA thioesters as substrates, without the involvement of ACPs (**Figure 1.8b**). Instead, a single homodimeric KS-like enzyme catalyzes the priming, iterative extension, and cyclization reactions. Examples of this type of PKS include the RppA synthase giving rise to flaviolin (**Figure 8b**), and DpgA from the biosynthetic pathway to 3,5-dighydroxyphenylglycine and the vancomycin family of antibiotics [77].

Although there are differences in the architecture and composition of the three types of PKS, the catalyzed chemistry is similar. All three types of PKS carry out the successive decarboxylative condensation of extender units by KS(-like) enzymes, using malonyl-CoA or its derivatives as building blocks. However, the type I modular PKSs, which can be likened to molecular scale assembly lines, are the most interesting in terms of potential applications, as there is a direct relationship between the architecture of the PKSs and the structure of their products. Indeed, the elegant co-linear biosynthetic logic of the modular type I PKSs has stimulated efforts over several decades to make polyketide derivatives of predictable structure for biological evaluation (**see section 1.7**).

1.6.2 Biosynthesis of polyketide synthase extender units

As mentioned earlier, PKS machinery accepts a variety of monomers as building blocks. Reflecting their roles in the biosynthesis, the starter unit is used to initiate polyketide synthesis, while extender units are used to elongate the chains. With an increasing number of starter units and extender units being identified and novel mechanisms for their incorporation being elucidated (see expert reviews by Moore and Hertweck 2002, Chan et al. 2009 and Ray and Moore 2016 [78-80]), they represent an important route for the introduction of structural diversity into polyketide scaffolds. As they are most relevant to the work described here, a brief treatment of extender unit biosynthesis is provided.

In general, extender units found in bacterial PKSs can be divided into two categories according to their mode of activation. Most commonly, extender units are activated as their CoA thioesters (e.g. malonyl-CoA and methylmalonyl-CoA). Alternatively, extender units are synthesized de novo on dedicated ACP domains; these include methoxymalonyl-ACP, hydroxymalonyl-ACP and aminomalonyl-ACP (**Figure 1.9**) [79].

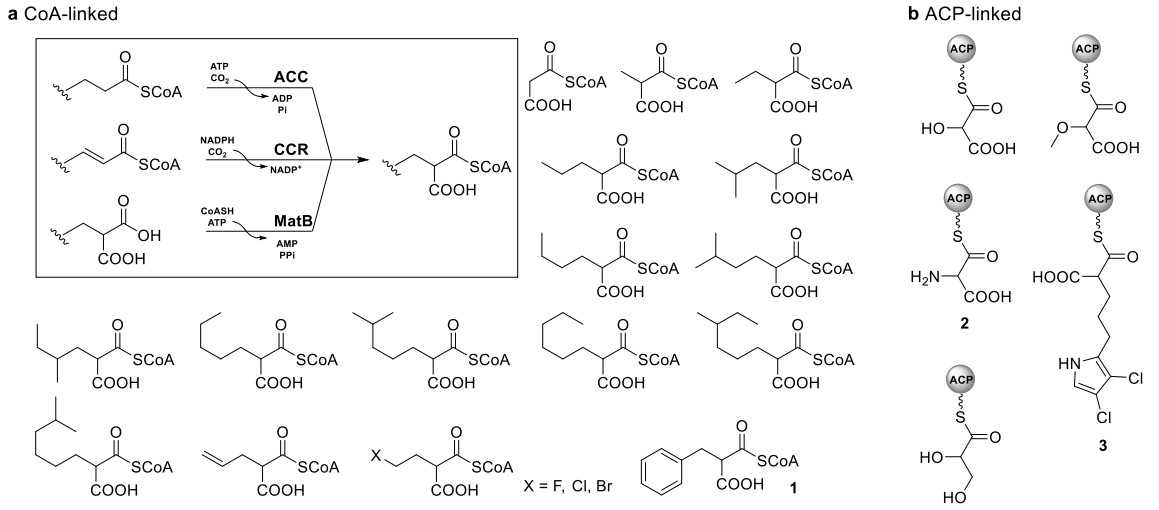


Figure 1.9 Structures of natural polyketide extender units. Benzylmalonyl-CoA (**1**), aminomalonyl-ACP (**2**), and dichloropyrrolymalonyl-ACP (**3**) are notably derived from amino acids. Adapted from reference [81].

1.6.2.1 CoA-linked extender units

Substrates for CoA-linked extender units mostly originate from fatty acid metabolites and arise via three main routes: 1) the α-carboxylation of acetyl-/propionyl-CoA by acetyl-CoA carboxylase (ACC) and propionyl-CoA carboxylase (PCC), respectively [82]; 2) ligation of malonate to CoA carried out by malonyl-CoA synthetases (MatB) [83]; and, 3) reductive carboxylation of α, β-unsaturated acyl-CoA by crotonyl-CoA carboxylase/reductase and its homologs (CCRs) [84].

1.6.2.1.1 ACC and PCC machinery

Malonyl-CoA and methylmalonyl-CoA, formed by ACC and PCC machinery respectively, are the two extender units most commonly used by modular PKSs. ACC and PCC are ATP- and biotin-dependent enzyme complexes and catalyze the carboxylation via a similar mechanism (**Figure 1.10a**). The two enzymes consist of three components: a biotin carboxylase (BC), a biotin carboxyl carrier protein (BCCP) and a carboxyl transferase (CT).

The BC component catalyzes the ATP-dependent activation of bicarbonate and the subsequent carboxylation of the N1 atom of the biotin prosthetic group of BCCP. Once the carboxyl-BCCP intermediate is formed, the CT activates the methyl group of acetyl-CoA for nucleophilic attack of the activated carboxyl group tethered to BCCP, resulting in the transfer of the carboxyl group onto acetyl-CoA, forming the malonyl-CoA [79] (**Figure 1.10b**). In *Streptomyces*, ACC and PCC share a common α-subunit, which consists of the BC and BCCP, but distinct β-subunit, which consists of the CT component and with a preference for either acetyl or propionyl-CoA as substrate.

The most intriguing of the recently characterized PKS extender units are 4-methylhexyl-/5-methylhexyl-/4-methylpentyl-/hexyl-CoA, a series of medium chain acyl-CoAs synthesized by a unique enzyme SamR0483, from the stambomycin biosynthetic pathway in *Streptomyces ambofaciens* ATCC23877 (presented in detail in **section 1.8**). Via the incorporation of isotope-labelled precursors, Ray and colleagues demonstrated that these atypical extender units originate from fatty acid metabolites [85]. Coupled with the X-ray crystal structure characterization of SamR0483 alone and in complex with its substrate, they further revealed the molecular basis for medium chain acyl-CoA recognition. Compared to the substrate binding pocket of the PCC β-subunit from *S. coelicolor* A3(2), SamR0483 contains more hydrophobic residues at the entrance of substrate binding pocket, contributing to its broad substrate tolerance, and allowing it to accommodate longer acyl chains than PCC [85]. Inspired by the intrinsic promiscuity of SamR0483, Ray et al. pursued a PDB approach to stambomycin diversification, providing alternative substrates to the strain bearing alkyne and azide groups. Consistent with SamR0483's substrate tolerance, this resulted in the corresponding analogues, albeit at yields only suitable for analysis by mass spectrometry. This finding suggested that SamR0483 might be a good candidate for development as a tool for PKS engineering. In this work, we carried out further experiments to exploit SamR0483, but also targeted it for inactivation in order to carry out mutasynthesis, which we anticipated might improve the yields of the desired derivatives (see **Chapter III)**.

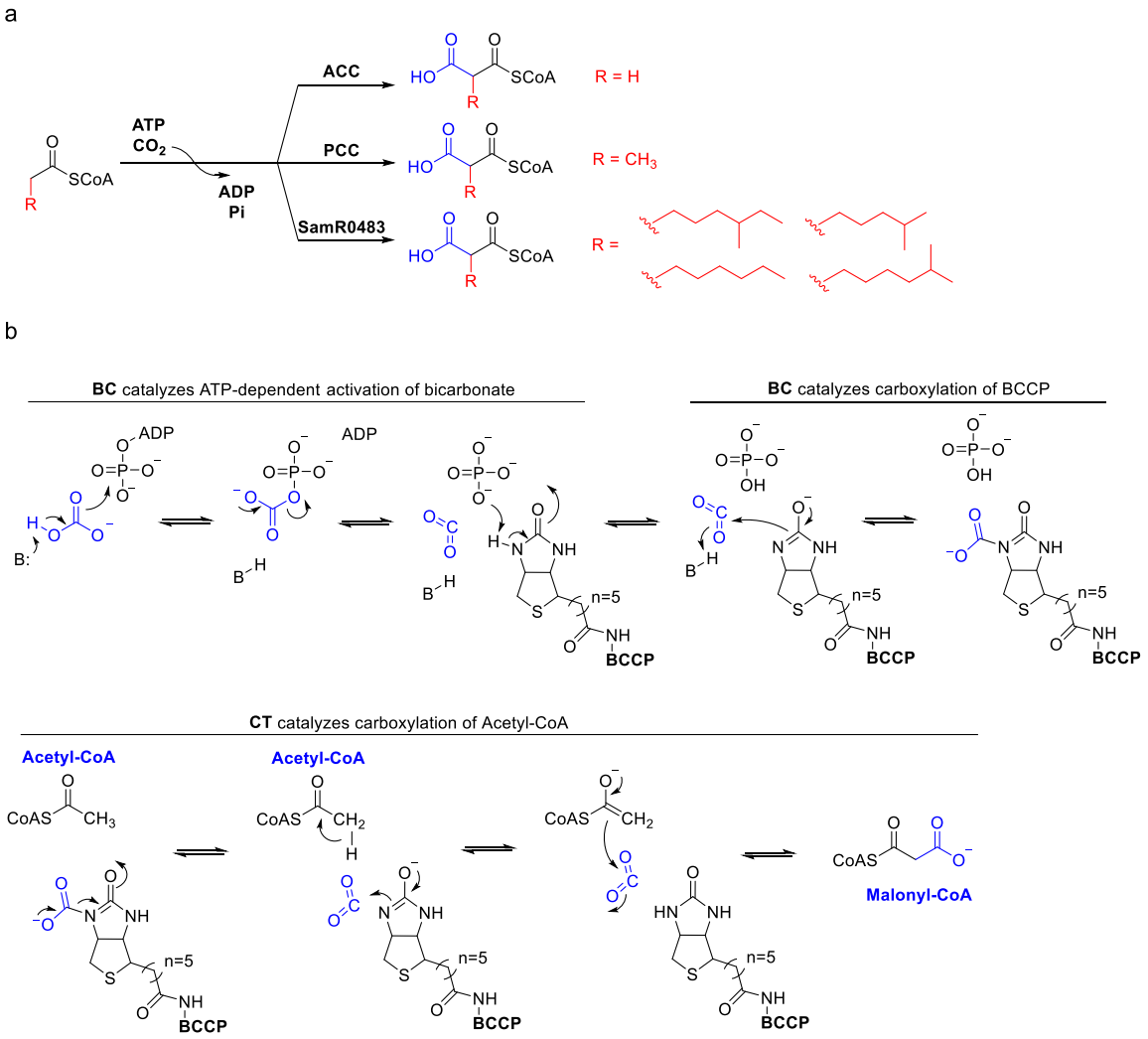


Figure 1.10 Biosynthesis of CoA-linked extender units by ACC/PCC/SamR0483-dependent pathways. **a**) ACC/PCC/SamR0483-catalyzed reactions. **b**) Mechanism of the ATP- and biotin-dependent reactions, exemplified by the ACC pathway. Adapted from reference [79]

1.6.2.1.2 MatB machinery

Unlike the ACC machinery, the malonyl-CoA can also be formed via ligation malonate and CoA by a malonyl-CoA synthetase in some bacteria (**Figure 1.11a**). The first bacterial enzyme capable of catalyzing this chemistry was identified from *Pseudomonas fluorescens*. Subsequently, a homologous enzyme, termed MatB, was identified from *Rhizobium trifolii*, in which exogenous malonate is imported *via* a membrane-bound dicarboxylate transporter protein and is directly activated to the CoA form by an ATP-dependent malonyl-CoA synthetase [86]. Further investigation of *R. trifolii* MatB indicated that it has broad

tolerance toward unnatural diacid substrates, providing an attractive route to expand the scope of precursors available for polyketide biosynthesis [87]. Further related work is discussed in **section 1.7.6**.

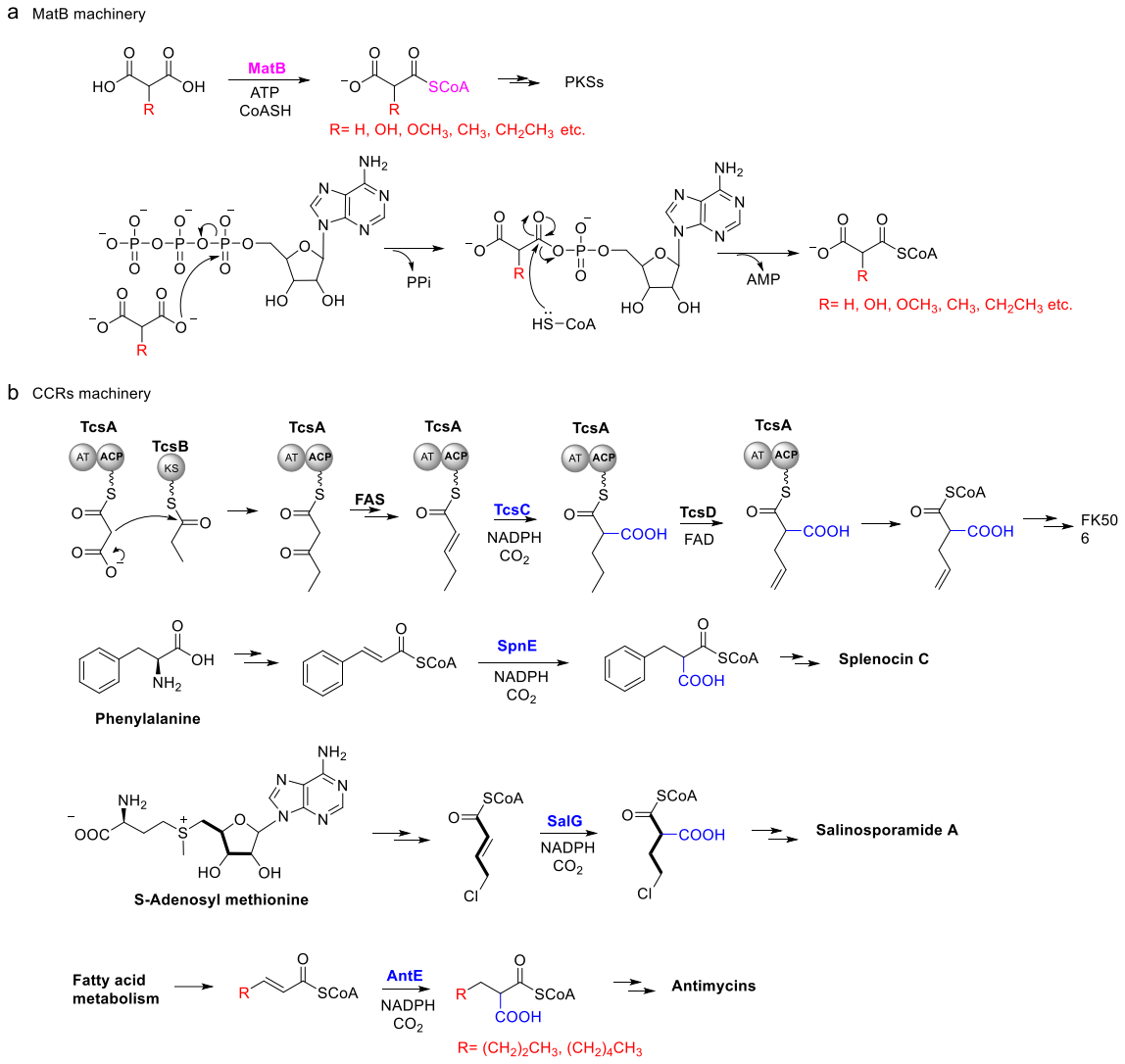


Figure 1.11 Biosynthesis of CoA-linked extender units by MatB and crotonyl-CoA reductase/carboxylase (CCR) pathways. a) Two-step reaction mechanism for MatB from *S. coelicolor* A3(2). In the adenylate-forming reaction, the ATP α-phosphate is attacked by a malonate carboxylate to form the adenylate intermediate and PPi. In the subsequent thioester forming reaction, CoA displaces AMP to produce the extender unit and AMP. **b**) Generation of malonyl-CoA derivatives as unusual PKS extender units via crotonyl-CoA reductase/carboxylase (CCR) machinery. All the extender units presented here are biosynthesized by the wild type MatB and CCRs. Exploiting these enzymes to extend the scope of extender unit biosynthesis is discussed in **section 1.7.6**.

1.6.2.1.3 CCR machinery

In addition to the above two pathways, CCRs are the third source for PKS extender units, and in particular unusual structures, including allylmalonyl-CoA (FK506), benzylmalonyl-COA (splenocins), and chloroethylmalonyl-CoA (salinosporamide A). In all of these cases, the final biosynthetic step is catalyzed by a homologue of CCR (**Figure 1.11b**), with the gene encoding CCR homologue clustered with the polyketide synthase genes. Given the pivotal role of CCRs in the biosynthesis of PKS monomers, their presence within a polyketide biosynthesis gene cluster is highly indicative of a distinct polyketide structure, and thus a useful marker for the potential priority of the cluster for genome mining. Recently, increasing numbers of atypical monomers have been unveiled. Taken together, these findings have inspired an approach based on CCRs (or engineered CCRs) as tools for incorporation of alternative buildings into polyketides (see **section 1.7.6** for further details).

1.6.2.2 ACP-linked extender units

Unlike the CoA-linked extender units, ACP-linked extender units are mainly derived from glycolate and amino acids. In order to be active, the acyl carrier protein needs to be modified post-translationally from its *apo*- to its *holo*-ACP form by a phosphopantetheinyl transferase (PPTase), which catalyzes the addition of a 4′-Ppant moiety derived from CoA to the conserved serine residue of the ACP. These ACP-linked extender units are unique to polyketide biosynthesis, with the genes coding for their biosynthesis found within or flanking the associated biosynthetic gene clusters. The first ACP-linked PKS extender unit to be reported was (2*R*)-methoxymalonyl-ACP in the biosynthesis of the macrolide FK520 from *Streptomyces hygroscopicus*, whose use results in incorporation of a methoxyacetyl group in the polyketide scaffold [88]. Insight into the assembly of this extender unit was gained by the sequencing and analysis of the associated gene cluster. Specifically, Wu and co-workers demonstrated that a subset of five genes encoding the enzymes FkbG, FkbH, FkbI, FkbJ (ACP) and FkbK were involved in the formation of this extender unit (**Figure 1.12a**). Firstly, FkbH, a phosphatase, covalently tethers a glycolytic pathway intermediate to the Ppant group of *holo*-FkbJ (an ACP), forming glyceryl-FkbJ. Next, FkbK, an NAD^+-dependent oxidase, oxidizes glyceryl-FkbJ to an aldehyde intermediate, followed by oxidation to hydroxymalonyl-FkbJ by the FAD-dependent FkbI. Finally, the SAM-

dependent methyltransferase FkbG, catalyzes the *O*-methylation of hydroxymalonyl-FkbJ to form (2*R*)-methoxymalonyl-FkbJ (**Figure 1.12a**) [88]. Further studies revealed that genes encoding homologues of these five enzymes are located in the BGCs for a number of other methoxyacetyl-containing PKs, including ansamitocin P-3, concanamycin A1, geldanamycin, herbimycin, oxazolomycin, spiramycin and tautomycin (**Figure 1.12a**) [89-92].

A variant of this pathway in which the final *O*-methyl transferase is absent gives rise to another extender unit, (2*R*)-hydroxylmalonyl-ACP. This is the case for the zwittermicin A BGC, which contains genes *zmaN, D, G, E*, encoding respectively homologues of FkbH, FkbJ, FkbK and FkbI, necessary to convert the glycolate to (2*R*)-hydroxylmalonyl-ACP (**Figure 1.12b**). Zwittermicin is constructed from an additional atypical extender unit, (*2S*)-aminomalonyl-ACP. An examination of the BGC revealed three additional genes *zmaH, I, J* encoding homologues of FkbJ, FkbI (FAD-dependent dehydrogenase) and a new functional A (adenylation) enzyme, ZmaJ. ZmaJ notably allows incorporation of L-serine instead of glycolate (**Figure 1.12c**), thus giving rise to a hybrid polyketide-nonribosomal peptide structure.

Another amino acid-derived building monomer particularly worth highlighting is dichloropyrrolyl-ACP, which is used mainly as a starter unit to prime NP biosynthesis. This biosynthetic strategy is exemplified by pyoluteorin, pyrrolomycin C, marinopyrrole A and chlorizidine A [80]. The most intriguing among these is the use of dichloropyrrolypropylmalonyl-ACP as extender unit by the chlorizidine A PKS, as it derives from a FAS pathway based on dichloropyrrolyl-ACP (**Figure 1.12d**) [93].

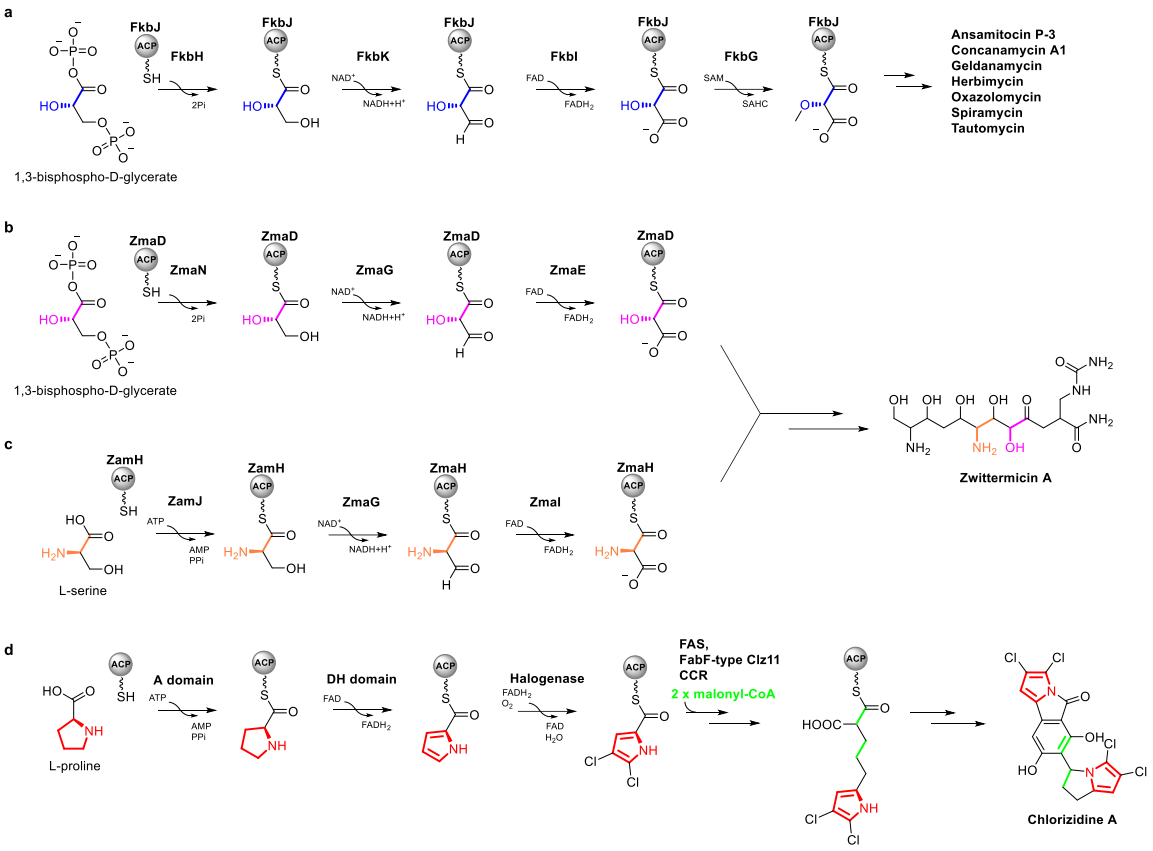


Figure 1.12 Biosynthetic pathway to ACP-tethered extender units and their corresponding polyketide products. a) In *S. hygroscopicus*, a five-gene set encodes the enzymes responsible for (2*R*)-methoxymalonyl-ACP in the FK520 biosynthetic pathway. Homologues of these genes are present in the BGCs for concanamycin A1, geldanamycin, herbimycin, oxazolomycin, spiramycin and tautomycin. **b**) Pathway to (2*R*)-hydroxymalonyl-ACP, for which the FkbG-homologue is absent. **c**) A novel pathway from serine to aminomalonyl-ACP involving an FkbI-homologue (ZmaI), FkbK-homologue (ZmaG), carrier protein ZmaH, and a special enzyme ZmaJ with specificity for L-serine. Both (2*R*)-hydroxymalonyl-ACP and aminomalonyl-ACP are utilized as extender units in zwittwermicin A biosynthesis. **d**) Biosynthesis of L-proline derived monomers dichloropyrrolyl-ACP and dichloropyrrolypropylmalonyl-ACP in the biosynthesis of chlorizidine A. The biosynthesis starts from the loading of L-proline onto an ACP by an adenylation (A) domain, followed by FAD-dependent oxidation (dehydrogenase), which converts the prolyl ring to a pyrrole. Subsequent halogenation by an $FADH_2$-dependent halogenase (Hal) then installs two halogens on the pyrrole ring, resulting in dichloropyrrolyl-ACP, which is used as a starter unit to prime the biosynthesis of chlorizidine A (top red pyrrole ring). Extension of dichloropyrrolyl-ACP with malonyl-CoA is then carried out by Clz11, a FabF-type enzyme, a FAS and a CCR in parallel, giving rise to dichloropyrrolypropylmalonyl-ACP, which is then incorporated into the chlorizidine A scaffold as an extender unit (bottom red pyrrole ring). Adapted from reference [79, 80].

1.7 Genetic engineering of modular PKSs

Modular polyketides play a pivotal role in NPs discovery, not only due to their enormous structural diversities and biological activities, but also the underlying biosynthetic logic. There has been longstanding interest in engineering these enzymes to produce novel polyketides in a predictable manner. Decades of work ranging from mutagenesis (amino acids substitution), deletion, insertion or replacement of domains, to exchange of domains, modules, or even entire subunits, have resulted in many new-to-nature polyketides, as well as providing important new insights into how modular PKSs function. In the following section, selected successful examples are reviewed, in order to illustrate strategies available for modular PKS engineering.

1.7.1 Domain modification

1.7.1.1 Tailoring domain KR-ER-DH modification

To date, the most frequent target of PKS synthetic biology has been individual domains or domain combinations. This type of modification allows access to the oxidation state at the β-keto position, inactivation, deletion, or insertion of tailoring domains. In fact, the first successful PKS engineering experiments targeted KR and ER domains of the DEBS system. Specifically, disabling ER_4 (ER domain of module 4; the same nomenclature is applied throughout) and KR_6, resulted in 6, 7-unsaturated and 3-keto-substitued 6-dEBs, respectively (**Figure 1.13** compound **1** and **2**) [94, 95]. However, a contemporaneous attempt to generate a DH_4 mutant to produce the full-length derivative failed, even though just a single active site histidine was altered [96]. Recently, KR, DH and ER domains have been targeted for site-directed inactivation in the monensin pathway, resulting in a library of 22 oxidized derivatives of premonensin (**Figure 1.14**).

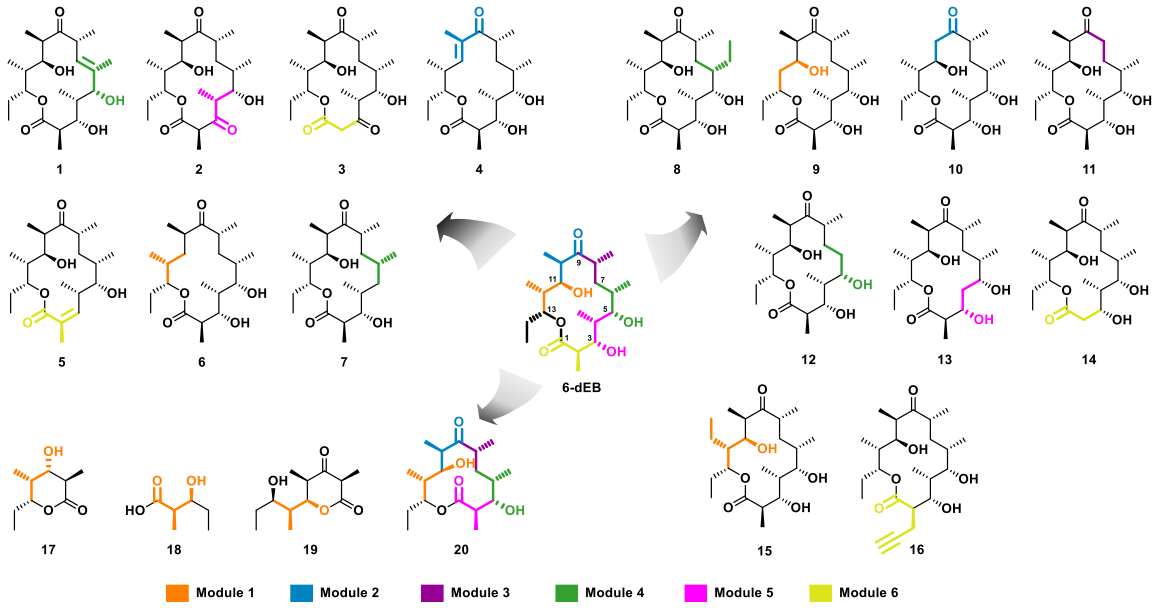


Figure 1.13 6-deoxyerythronolide (6-dEB) analogues generated by DEBS domain manipulation. The introduced modification include inactivation, deletion, and insertion of the tailoring domains KR, DH and ER (compounds **1–7**), mutation, *trans*-AT complementation and replacement of AT domains (compounds **8**–**16**), and relocation of the TE domain (compounds **17**–**20**). The modified moieties in the analogues relative to 6-dEB have been highlighted in accordance with the color of their respective modules.

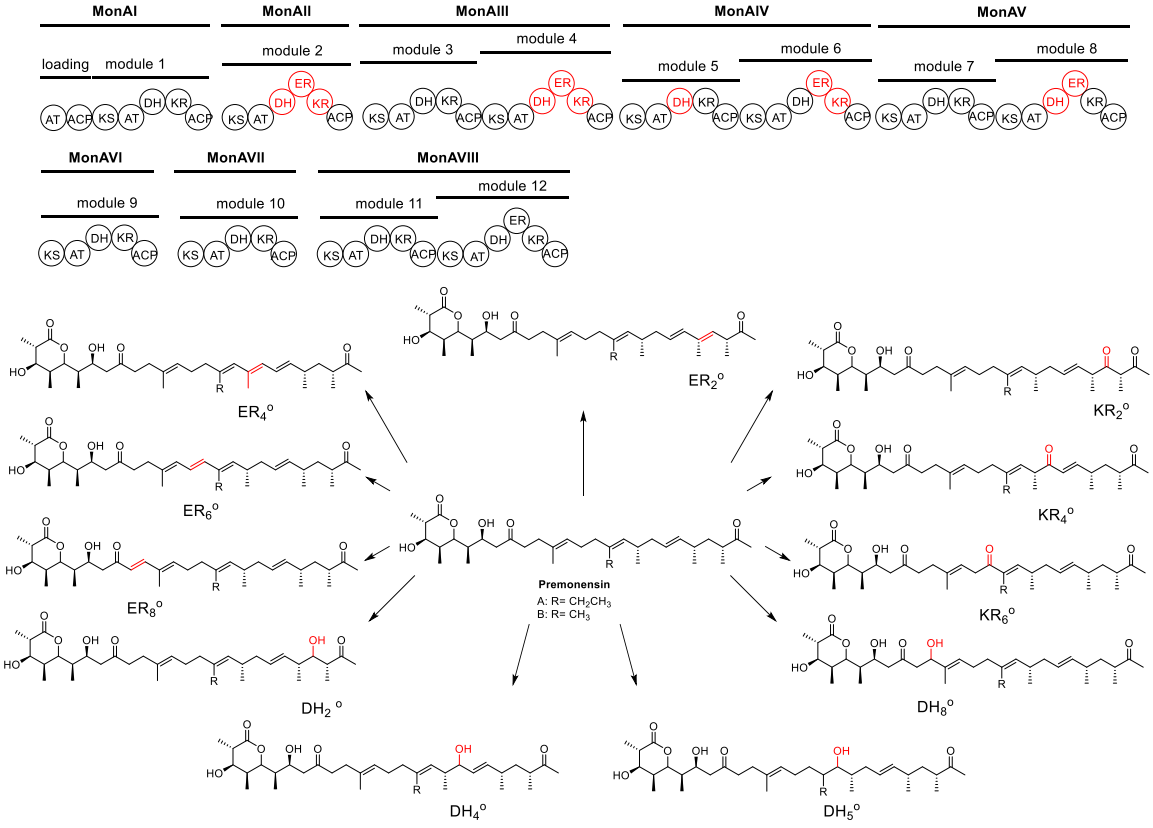


Figure 1.14 Set of derivatives of premonensin generated by site-directed inactivation of tailoring domains KR, DH and ER. The PKS organization of monensin is shown above, in which the domain targeted in each case is indicated in red, the modified and introduced moieties are color-coded in red. Adapted from reference [63].

Deletion of KR, DH and ER domains is an alternative strategy for remodeling the oxidation state of β-keto. For instance, when the entire DEBS KR_6 domain was deleted (replacement of the 447 amino acid regions between AT_6 and ACP_6 by 18 amino acids linker), the desired 3-keto-substituted 6-dEB was detected, as with the KR_6-inactivation mutant mentioned above. However, an unexpected product, 2-demethylated 6-dEB, was also produced, probably due to the mis-selection of an malonate extender unit rather than a methylmalonate by module 6 (**Figure 1.13**, compound **3)** [95].

Furthermore, individual KR domains have also been exchanged for reductive domain sets (e.g. DH-KR or DH-ER-KR). Such a 'gain-of-function' experiment was used to convert a hydroxyl group to a double bond by exchanging KR_2 from module 4 of the rapamycin (RAPS) PKS for a DH-KR didomain, resulting in 10, 11-dehydrated 6-dEB (**Figure 1.13**, compound **4**). Similarly, the exchange of KR_6 domain for this set of domains resulted in 2, 3-dehydrated 6-dEB (**Figure 1.13**, compound **5**). In addition, 11-dehydroxylated and 5-dehydroxylated 6-dEBs (**Figure 1.13**, compounds **6** and **7**) were identified when KR_2 and KR_5 were replaced with the DH-ER-KR domains from module 1 of RAPS PKS, respectively. However, no predicted 3-dehydroxyl product was detected in the case of the KR_6 replacement, but the 3-keto-substituted and 2,3-dehydrated 6-dEBs were found instead, showing that the DH and DH-ER domains sometimes failed to act (**Figure 1.13**, compounds **1** and **5**). Therefore, even though the manipulation of the KR, DH, and ER domains has been fruitful, some experiments fail, potentially due to alterations in PKS architecture which impact the ability of certain domains to function in each cycle, or to incompatibilities in substrate specificity.

1.7.1.2 AT modification

The AT domain acts as a 'gate keeper' in recruiting building blocks for incorporation into polyketide scaffolds, and thus, manipulating AT domains is a promising strategy for introducing distinct side chains into the polyketide skeleton. In DEBS, the AT domains of

all the extension modules are specific for methylmalonyl-CoA. Swapping the AT_4 domain by an ethylmalonyl-CoA specific AT_5 from the niddamycin PKS assembly line, for example, resulted in a new erythronolide bearing a 6-ethyl group (**Figure 1.13**, compound **8**) [97]. Similarly, replacement of $AT_{1–6}$ with AT domains specific for malonyl-CoA in other PKS systems, respectively, gave rise to a series of demethylated 6-dEBs (**Figure 1.13**, compounds **9**−**14**) [98-101].

Another notable strategy is termed '*trans*-AT complementation'. This strategy refers to the use of an exogenous *trans*-acting AT for loading extender units onto a modular polyketide synthase, in which an AT domain had been inactivated by substituting of the active site Ser residue with Ala. A notable example of this strategy was the use of the ethylmalonate specific AT_5 from the kirromycin *trans*-AT PKS system to rescue the inactivated AT_1 of DEBS. This experiment gave rise to the expected 6-dEB incorporating an ethyl group at the C-12 position (**Figure 1.13**, compound **15**) [102].

Alternatively, the substrate profile of an AT domain can be altered by targeting its putative specificity-determining residues. Initially, sequence alignment identified consensus motifs correlating with the AT substrate specificity: 'HAFH motif' in malonyl-CoA recruiting AT domain, and 'YASH motif' in ATs specific for methylmalonyl-CoA [103]. Subsequently, additional motifs were identified at the same location, including 'HASH' present in the AT_3 domain of the epothilone PKS which accepts both malonyl- and methylmalonyl-CoA as substrates, and the 'CPTH motif' of the AT_4 domain in FK506 biosynthesis which accepts allylmalonyl-CoA. Despite these attractive correlations between sequence and specificity, mutation of this motif did not engender a complete shift in substrate choice, and was associated with reduced activity *in vivo* [104]. It is also notable that a single point mutation (from valine to alanine) in DEBS AT_6, was sufficient to shift specificity from methylmalonate to non-native 2-propargylmalonate, resulting in 2-propargyl-6dEB although 6-dEB was also present. (**Figure 1.13**, compound **16**) [105]. Taken together, these experiments that the current information available on AT specificity do not permit the complete redirection of substrate specificity. Nonetheless, recent studies *in vitro* on a series of AT domains based on solved crystal structures produced a dramatic

shift in k_{cat}/K_M towards unnatural substrates, but the utility of these mutations in vivo has yet to be demonstrated [106].

1.7.1.3 TE relocation

Another productive strategy has been to relocate chain-terminating TE domains to the C-terminus of various ACPs in order to effect premature release of polyketide chains. This approach was first applied to the DEBS system, when the TE from the end of DEBS 3 was grafted onto the C-terminus of DEBS 1, resulting in the predicted triketide lactone (**Figure 1.13**, compound **17**) [107]. The resulting "DEBS 1-TE" mini PKS system has frequently served as a model for understanding a number of mechanistic issues in polyketide biosynthesis [63]. Similarly, when the TE domain was moved to the C-terminus of modules 1, 3 and 5, the engineered systems gave rise, respectively, to the diketide, tetraketide and hexaketide products (**Figure 1.13**, compounds **18**−**20**) [63]. Evidently, the DEBS TE domains tolerates a range of different polyketide lengths. It was subsequently relocated into other systems including the spinosyn, monensin and lipomycin PKSs, catalyzing successful chain release in each case [63]. In comparison with other domain-level manipulation strategies, TE relocation is more likely to give the expected products at a relatively high titer (for example, the pentaketide of the monensin intermediate was produced at 44 mg/L) (**Figure 1.15**) [108].

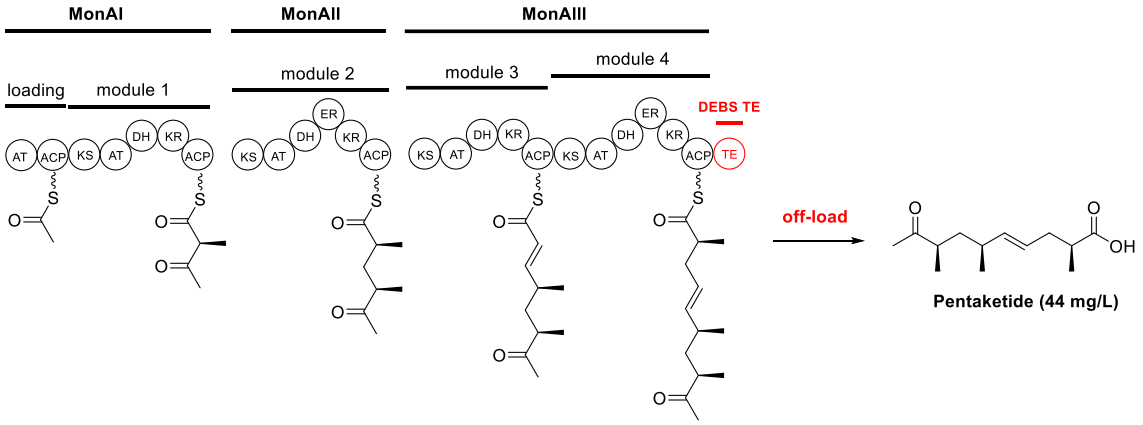


Figure 1.15 Truncated pentaketide generated from the monensin PKS by grafting of the DEBS TE domain to the C-terminal end of extension module 4.

1.7.2 Module swapping

PKSs have also been manipulated at the module level. For example, when the loading module of DEBS was swapped with that from the avermectin PKS (AVES) system, novel erythromycin derivatives were generated bearing the branched 2-*iso*-propyl and 2-*sec*-butyl starter units characteristic of avermectin at the C-13 position (**Figure 1.16**) [109]. Subsequently, the unusually broad specificity of the AVES loading module was exploited to generate a large range of 6-dEB derivatives by PDB (by feeding various fatty acids). (**Figure 1.16**) [109].

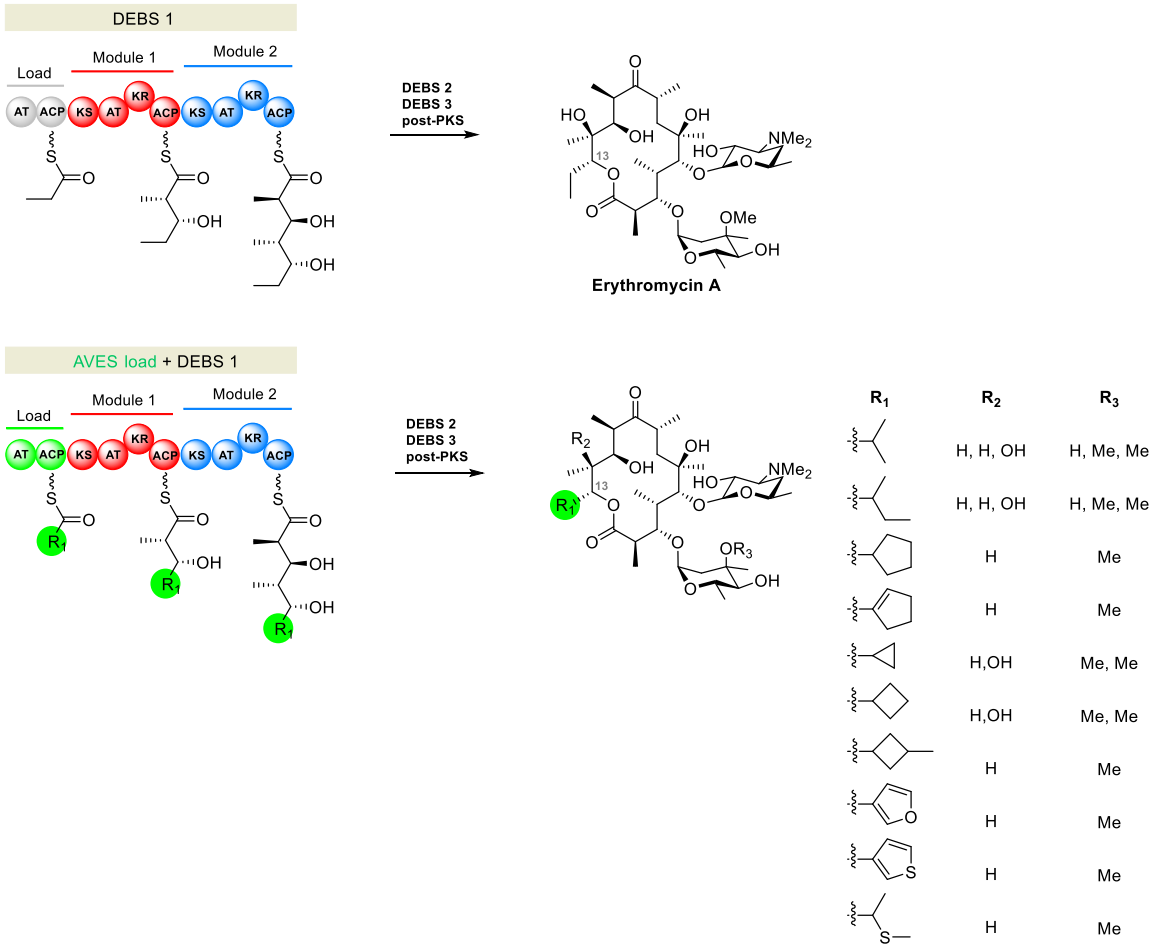


Figure 1.16 The products of a hybrid modular DEBS incorporating an altered loading module. The loading module for DEBS natively accepts propionyl-CoA. Replacement with the loading module from the AVES PKS system altered loading specificity, resulting in erythromycin analogues. Adapted from reference [63].

In addition to the loading module swapping, chain extension modules of some PKSs have been modified. By using the DEBS 1-TE mini-PKS system, new triketide lactones were obtained when the original module 2 was substituted by downstream modules 3 and 6, as well as module 5 from the rifamycin (RIF) PKSs, by conserving the natural intermodular linker between modules 1 and 2 (**Figure 1.17b, c, d**) [63]. As an extension to this approach, by appending the ^{C}DD from DEBS module 2 to the end of RIF module 5, the recombinant DEBS M1-RIF M5 protein was induced to communicate with the downstream DEBS 2 and 3, resulting in full length 6-dEB (**Figure 1.17e**). Notably, this work laid the foundation for synthetic biology of polyketides utilizing modules, not domains, as the fundamental unit.

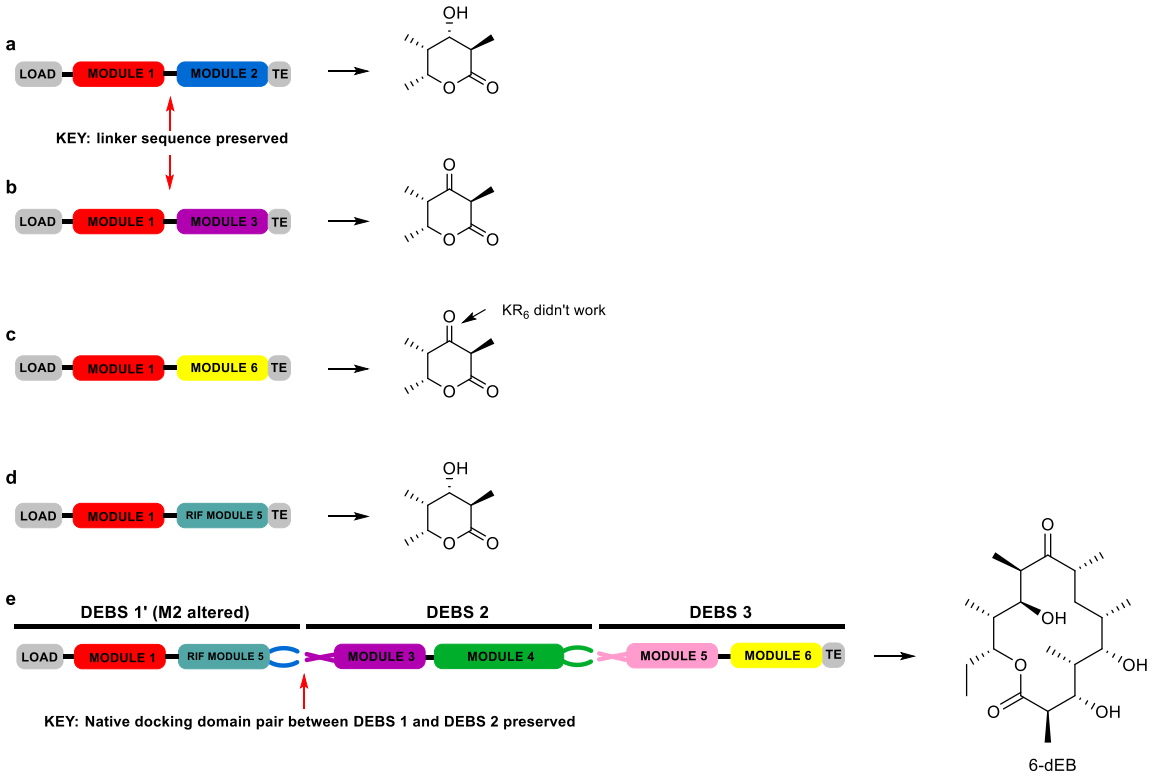


Figure 1.17 Whole module swapping as an engineering approach. a) DEBS 1-TE mini PKS model system and its triketide lactone product. **b, c)** Variants of DEBS 1-TE containing heterologous module 2s, and their respective triketide lactone products. Communication between module 1 of DEBS and two non-cognate modules from the DEBS system (3 and 6) was achieved by preserving the native intermodular linker between modules 1 and 2. **d)** The same strategy resulted in successful cross-talk between DEBS module 1 and module 5 from the rifamycin (RIF) PKS. **e)** By preserving the native docking interaction between subunits 1 and 2,

the full length 6-dEB product was then obtained from the RIF module 5 construct shown in **d**), in the presence of downstream subunits DEBS 2 and DEBS 3. Adapted from reference [63].

An attempt was made to insert modules 2 and 5 from the RAPS PKS in between modules 1 and 2 in the DEBS 1-TE system, with the aim of adding an extra round of chain extension. However, the major product was not the expected tetraketide, but a triketide (**Figure 1.18a**) [63]. An explanation given for this phenomenon was the 'skipping' of the interpolated modules (**Figure 1.18b**). According to this model, the polyketide chain was handed off directly from ACP_1 (DEBS) to ACP_2 (RAPS) instead of to KS_2 (RAPS), due to poor substrate recognition by KS_2 (RAPS). This observation suggested that the native ACP_n-KS_{n+1} interface should be maintained to preserve both the protein:protein interactions and substrate specificity underlying efficient transacylation [63]. The importance of such ACP/KS interactions for chain transfer was reinforced by in several subsequent studies [110, 111] (see **sections 1.7.3** and **1.7.4**).

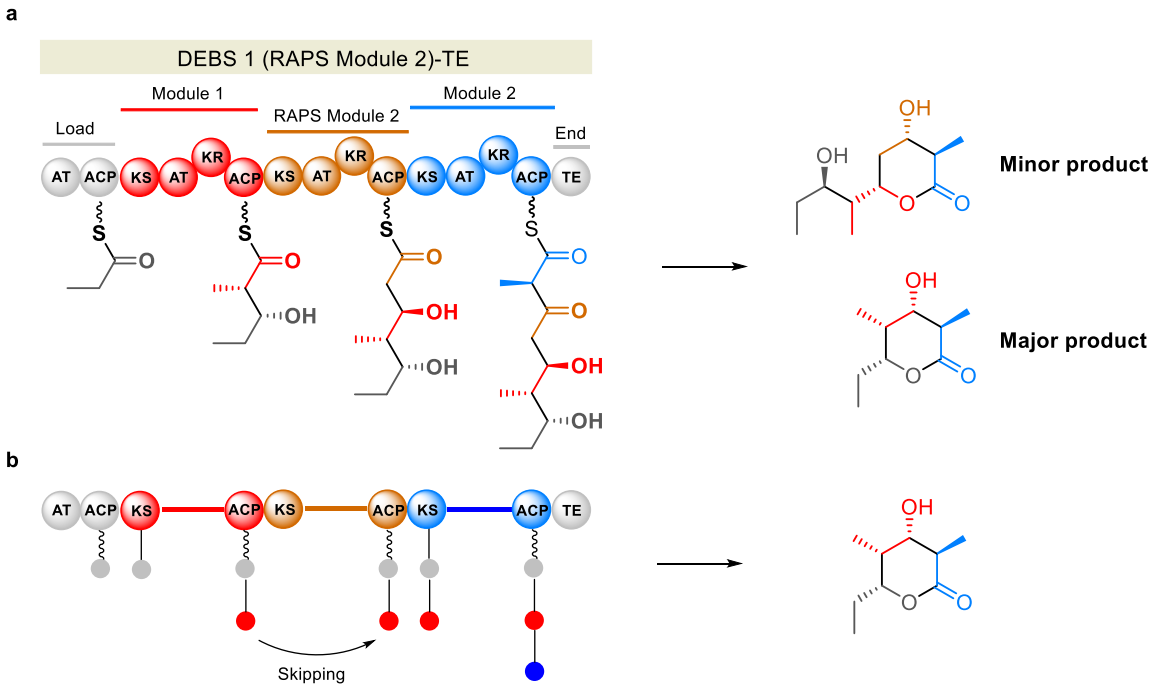


Figure 1.18 Attempted addition of a module to the DEBS system. a) RAPS module 2 was inserted at an internal position within DEBS 1, between modules 1 and 2. Although the predicted tetraketide was observed in minor amounts, the major product was triketide lactone, arising from skipping of the interpolated module. **b**) A separate study elucidated the molecular basis for the by-passing event, showing that the intermediate undergoes direct ACP-to-ACP transfer involving the inserted module. Adapted from reference [63].

1.7.3 Re-engineering PKS subunits

Another approach in recent PKS engineering is to use the subunit as the basic unit [112]. As mentioned in **sections 1.7.1** and **1.7.2**, changes to domains/modules within the PKS subunits may cause the PKS complex to fold incorrectly, leading to loss of activity or to modified chains that cannot be recognized as substrates by subsequent subunits. In contrast, replacement of a gene encoding the subunit itself should be a more reliable strategy that preserves the native relationships between the catalytic domains.

Underlying this type of strategy was the discovery of regions at the C- and N-terminal ends of adjacent PKS polypeptides, termed docking domains (DDs), which mediate the interactions between consecutive subunits [113], although the flanking ACP and KS domains also contribute to the interfaces, in *cis*-AT PKSs at least [110]. Notably, Khosla and co-workers showed that certain DDs are modular regions, which can be substituted by other matched pairs without negatively impacting affecting the activity of individual modules/subunits [114, 115]. Subsequently, bioinformatics analysis identified the first three distinct sequence classes of docking domains in *cis*-AT PKS systems: class 1a, 1b and class 2. High-resolution structural information is now available for all of these DDs (**Figure 1.19**) [55, 116-118], showing that the DD classes adopt distinct architectures, as well as helping to identify their boundaries and highlighting residues which might be targeted in attempts to alter interaction specificity.

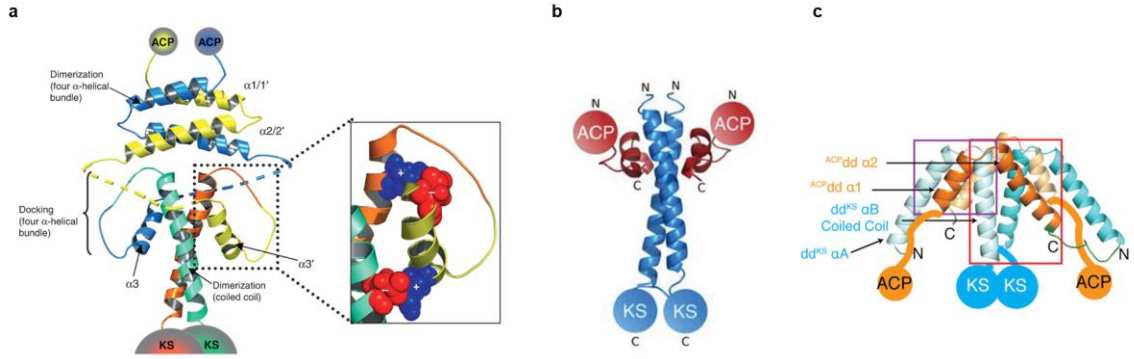


Figure 1.19 Structures of docking domains from modular *cis*-AT PKSs: Class 1a (PDB 1PZQ, 1PZR), Class 1b (PDB 3F5H) and Class 2 (PDB 4MYY). **a**) NMR structure of covalently fused DD complex, comprising the C-terminal DD (^{C}DD) of DEBS 2 and the N-terminal DD (^{N}DD) of DEBS 3. The overall structure contains two four-α-helix bundles [55]. The first, an intertwined bundle, is formed by the first two α-helices of the C-terminal DD (shown in blue and yellow). The second four-helix bundle is formed by the

third α-helix of the C-terminal DD which wraps around the coiled coil formed by the N-terminal docking domain (shown in green and orange). Residues involved in docking interface are highlighted in the box, and include F67 and L70 helix α3′ and Y96′, L97, T100′, L104 on helix α4, α4′ (hydrophobic interactions), as well as D64′ on helix α3′ and K92′ on helix α4′ (electrostatic interactions) [55]. **b**) X-ray structure of a covalent complex of the ^{C}DD of PikAIII and the ^{N}DD of PikAIV. The overall architecture is similar to that of the DEBS 2/DEBS 3 docking domain model, but the ^{C}DD is shorter than its DEBS counterpart, and the precise hydrophobic and polar interactions at the interface differ [117]. **c**) X-ray crystal structure of a covalent complex of class 2 docking domains. The ^{C}DD does not incorporate a dimerization region. Instead, its two α-helices associate with two α-helices provided by the N-terminal docking domain to form an overall 8 α-helical bundle [118]. Adapted from reference [63].

In nature, the modular PKS systems of DEBS, Pik and Ole produce the 14-membered macrolides 6-dEB, narbonolide and 8, 8a-deoxyoleandolide, respectively. Each PKS complex consist of six modules, and the amino acid sequences of modules 5 and 6 of Pik show 42% identities to that of DEBS and Ole (the alignment was carried out by considering Pik modules 5 (subunit PikAIII) and 6 (PikAIV) as a single subunit, such as present in DEBS and Ole). Additionally, the biosynthetic intermediates for modules 5 and 6 are similar among the PKSs. These similarities thus prompted an effort to investigate the cross-functionality of the various subunits. To this end, hybrid PKSs were constructed in the heterologous host *Streptomyces lividans*, in which the first two subunits from Pik were combined with DEBS 3 (containing modules 5 and 6) or the third subunit (containing modules 5 and 6) of the Ole assembly line. The chimeric PKSs produced the expected 14-membered lactones, showing that the native DDs of each subunit were able to communicate with the heterologous partners (**Figure 1.20**) [119].

Analogous set of experiments were also carried out between avermectin and milbemycin PKSs. Avermectin and milbemycin are 16-membered macrolides with insecticidal activity, but milbemycin shows higher activity against some parasites and lower toxicity than avermectin. The structural difference between avermectin and milbemycin occur at C-13, C-22, 23 and C-25 (**Figure 1.21**). Based on the structure and pharmacological property comparison among avermectin, milbemycin and ivermectin (a derivative of avermectin obtained by semi-synthesis), the single bond across C-22, 23 was thought to be the preferred structure in terms of lower toxicity, and the functionality at C-13 was thought to be crucial for antiparasitic activity. Hence, on the basis of the inherent

similarity of their biosynthetic pathways, the subunit AveA1 was exchanged with that of MilA1, resulting a hybrid Mil/Ave PKS, and successfully affording two novel metabolites called tenvermectins at high yields (3400 mg/L) in an industrial avermectin producing strain. Insecticidal activity tests showed that the tenvermectins indeed exhibit enhanced activity relative to avermectin (**Figure 1.21**) [120]. These results showcase the inherent ability of evolutionary-related subunits to interact productively, suggesting a feasible strategy towards altering or improving the biological properties of certain polyketides (i.e. those for which closely-related systems exist and have been characterized).

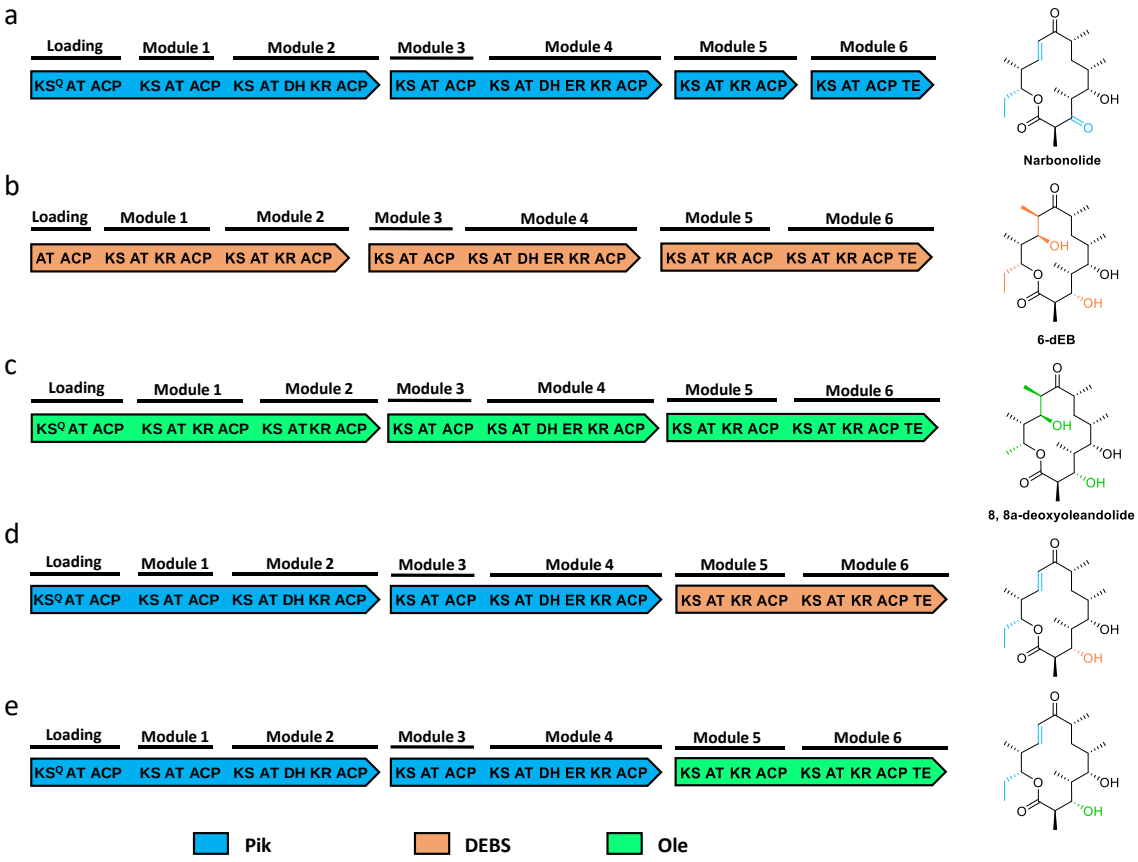


Figure 1.20 Modular arrangement of the native PKSs that produce narbonolide (a), 6-dEB (b) and 8, 8a-deoxyleandolide (c), and the whole subunit swapped chimeric PKSs (d and e) that produce hybrid 14-membered macrolides. The main differences between the structure of these macrolides are highlighted in a manner consistent with the color of their respective assembly lines. KS^Q indicates a KS domain whose catalytic cysteine is replaced by a glutamine residue. Adapted from reference [63].

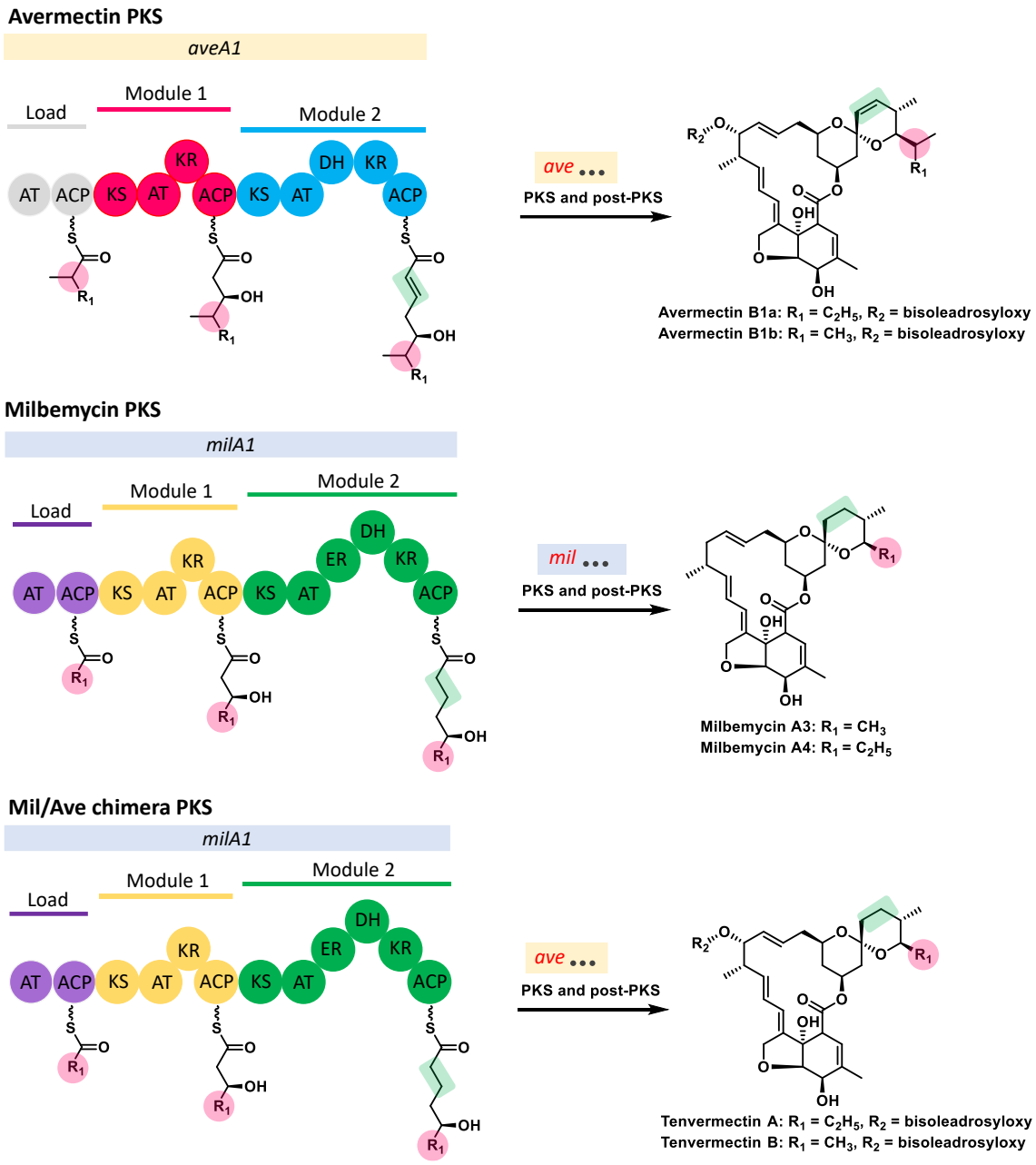


Figure 1.21 Generation of tenvermectins by exchanging the subunit AveA1 for MilA1. The differences between the parent avermectin and milbemycin structures are shown with highlighted colors, as well as the structure of the hybrid tenvermectins [120].

On the other hand, PKS engineering based on subunits is not intrinsically limited to closely-related systems. In such cases, however, it cannot be assumed that the domains/DDs at the novel junctions will be compatible, and thus additional interface engineering may be necessary. This approach is exemplified by the work of Watanabe and

co-workers. Specifically, they created a two-plasmid system to co-express DEBS 1 and a panel of acceptor cassettes, each cassette being composed of a single module with the DEBS TE fused to its C-terminus. In addition, to facilitate chain transfer to the acceptor cassette from DEBS 1, the N-terminal docking domain of DEBS 2 (called M3N) was appended at its N-terminus. Using this system, successful communication was observed between DEBS 1 and DEBS modules 3 and 6, as well as two other heterologous modules, RIF module 5 and Pik module 6 (**Figure 1.22a**) [121]. Similarly, interaction between DEBS 2 and two non-native modules, DEBS module 6 and RIF module 5, was observed when the N-terminal docking domain of DEBS 3 (M5N) was appended to the acceptor modules, resulting in the expected hexaketide (**Figure 1.22b**) [121].

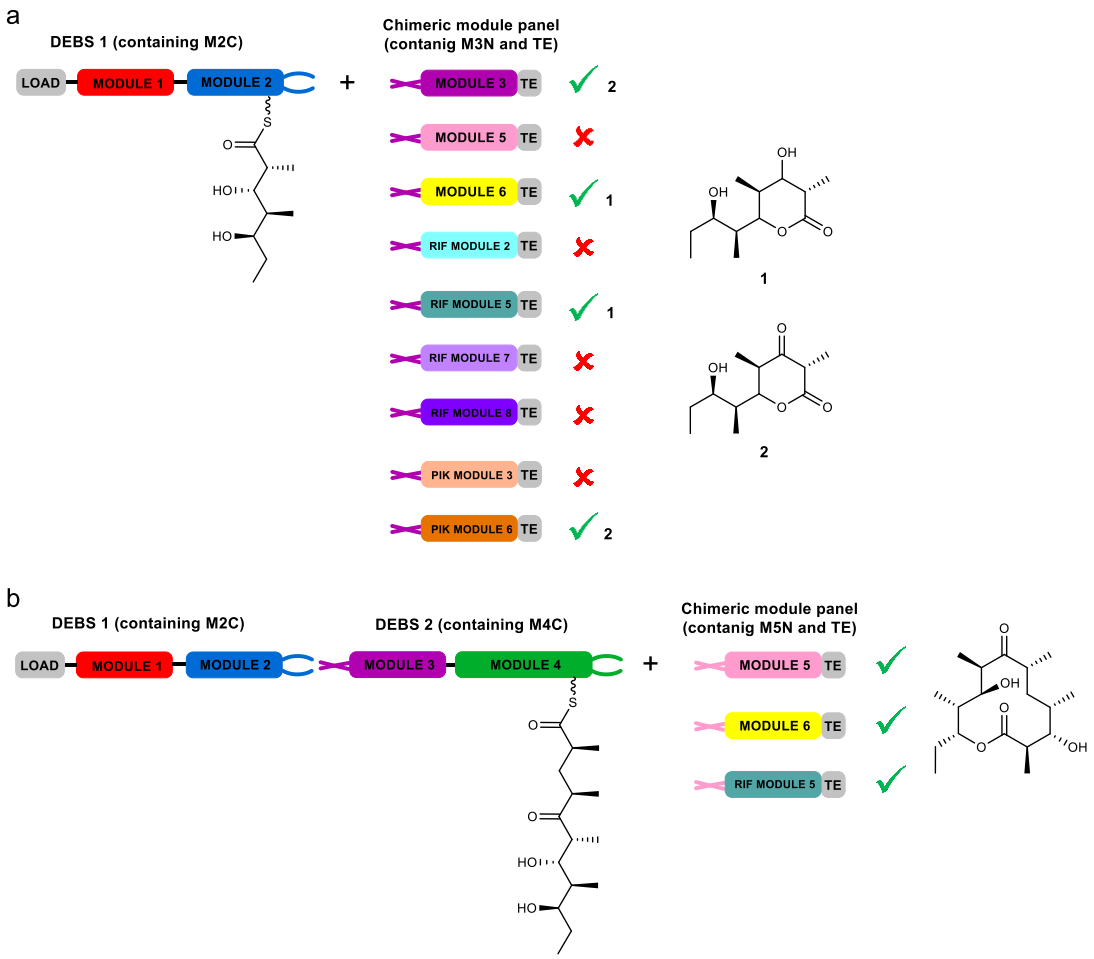


Figure 1.22 Scheme showing the strategy employed by Watanabe, *et al.* [121]**: use of docking domain engineering to facilitate chain transfer between non-native PKS partners.** The two-plasmid co-expression system furnished donor protein bearing chain extension intermediate, which was then combined

with various modules as acceptors. Communication between the two modules was facilitated by matched pairs of ^{C}DDs/^{N}DDs, resulting in some cases in the expected lactones. The failure of certain constructs in these experiments was attributed to the substrate selectivity of the acceptor modules, and in particular the KS domains. M2C and M4C refer to the native ^{C}DDs of modules 2 and 4. M3N and M5N refer to the native ^{N}DDs of modules 3 and 5.

Although a number of these experiments were successful, a sizeable proportion of acceptor cassettes in **Figure 1.22a** failed to process the chain extension intermediates despite the presence of compatible docking domains – a result attributed at least in part to poor tolerance by the downstream KS domain. To provide direct support for this hypothesis, module 2 in the donor protein DEBS 1 was substituted by Pik module 2. This construction resulted in an acyl-enzyme intermediate of alternative structure, incorporating α-C=C-β instead of α-C(CH_3)-C(OH)-β − thus allowing a direct test of the role played by substrate specificity in processing by the downstream module. Co-expression of this modified DEBS 1 with both DEBS and Pik module 3 resulted in the expected product in the case of Pik module 3 (although it was not active with native DEBS 1), but not in the case of DEBS module 3 (**Figure 1.23**) [121]. This result is consistent with an important contribution of substrate specificity to the activity of hybrid PKS systems, particularly with regards to the functionality at the α- and β- positions. In conclusion, a functional combination of PKS subunits requires not only an understanding of the factors mediating efficient protein-protein interactions, but also in certain cases, consideration of the substrate specificity of the downstream KS domain.

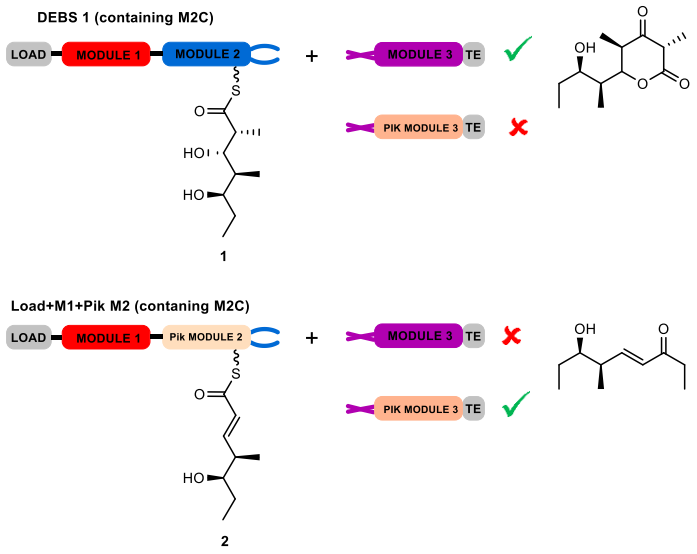


Figure 1.23 Assay for tolerance to variation at the substrate α and β position. Compounds 1 and 2 are the native substrate for DEBS module 3 and Pik module 3, respectively. This work showed that the stereochemistry at the α and β positions is crucial for recognition downstream proteins. Adapted from reference [121].

Furthermore, as mentioned in the section on module swapping, maintenance of cognate ACP_n-KS_{n+1} recognition (i.e. protein-protein interactions beyond the DDs) can also be important for efficient translocation of chain extension intermediate. For example, research by Chandran and co-workers showed that initially non-productive hybrid polyketide synthases could be rescued by re-establishing the original ACP_n/KS_{n+1} interface [122]. In recent follow-up work, Klaus and co-workers designed a set of experiments to elucidate the relative contributions of ACP_n/KS_{n+1} and enzyme-substrate interactions to turnover rates in chimeric PKSs (**Figure 1.24**). In their *in vitro* experiments, the first polypeptide of DEBS was split into three parts and expressed as three individual proteins: 1) the loading module (LM) was expressed with the C-terminal docking domain of module 4 (M4C) at its C-terminus; 2) DEBS module 1 was expressed with M5N (N-terminal DD of module 5, the native partner of M4C) at its N-terminus and M2C at its C-terminus; and, 3) DEBS module 2 was expressed with M3N at its N-terminus and TE at its C-terminus. In addition, a panel of modules originating from different PKSs was constructed and tested in place of DEBS module 2, to evaluate their ability to accept the same diketide substrate from the bimodular donor system (LM+DEBS M1). The obtained data revealed that the differences in turnover rates (10~100-fold less relative to the reference using native DEBS M2) mainly

reflected the extent of compatibility at the chimeric ACP_n/KS_{n+1} interface. Indeed, when the KR domain in DEBS M1 was swapped for that of DEBS M2, with the result that the stereochemistry of the intermediates was altered, the turnover rate to product fell by only a factor of 7-fold. The overall conclusion from this work was that protein-protein interactions, not substrate recognition, dominates the turnover of chimeric polyketide synthases [111].

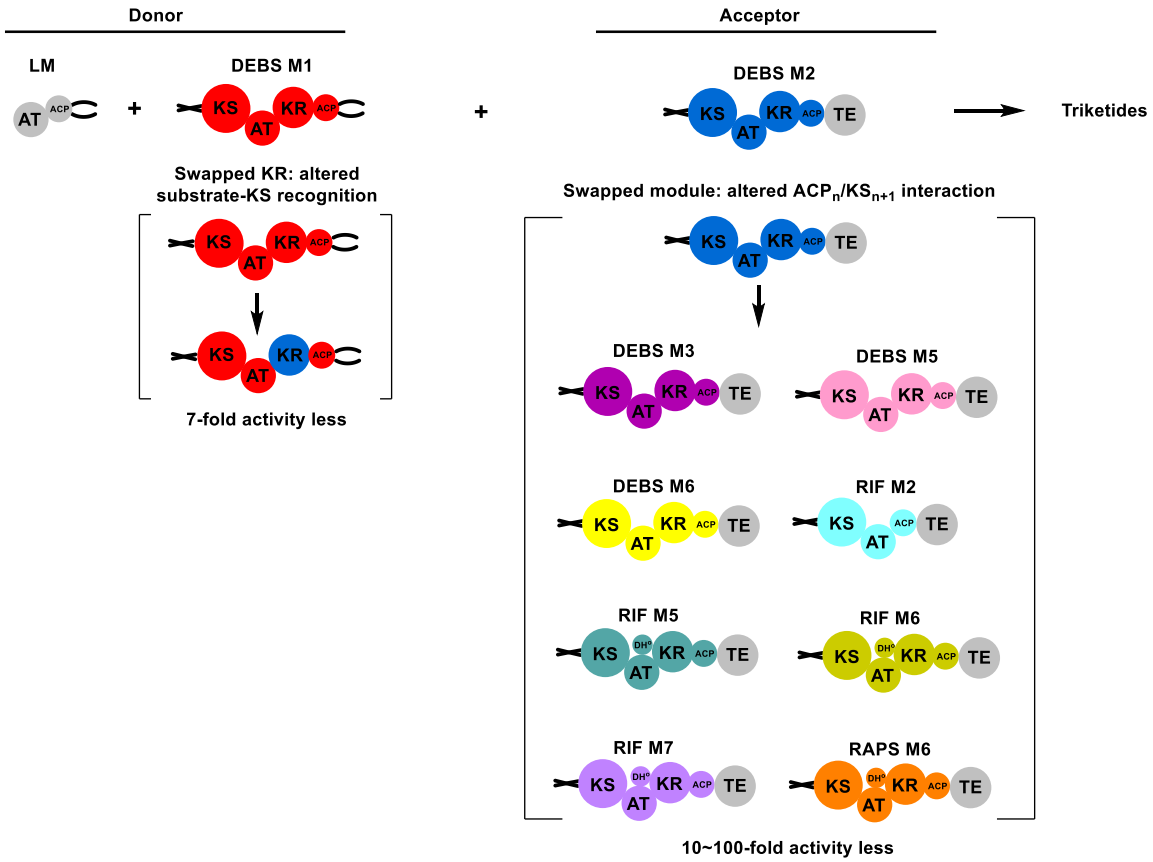


Figure 1.24 Quantitative comparison of the effect of protein-protein interactions and substrate recognition in chimeric bimodular PKSs. Adapted from reference [111].

1.7.4 Evolution-inspired genetic engineering

The rational approaches to PKS engineering discussed previously have achieved some success, but even functional hybrids were for the most part substantially compromised relative to the parental systems. To improve our ability to carry out PKS synthetic biology, researchers have increasing looked to the natural processes of PKS evolution for guidance.

PKS modules have traditionally been defined as contain the following domains: KS-AT-(DH-ER-KR)-ACP, which correspond to the domains required to catalyze a round of chain extension and any associated tailoring of the newly-incorporated building block. This domain set is also an architectural and genetic unit (e.g. PikAIII, **Figure 1.25**, and the many mono-module subunits that exist), and this domain order is conserved in mammalian fatty acid synthases. Nonetheless, Zhang and co-workers recently discovered that the KS domains of certain PKSs co-migrate during assembly line evolution with the upstream ACP domain and β-keto processing domains (DH-KR-ER) [123]. They observed this by comparing four closely-related, gigantic PKSs, which produce aminopolyols: mediomycin, neomediomycin, ECO-02301 and neotetrafibricin. Bioinformatic analysis revealed the relationships between equivalent domains within the four assembly lines, prompting the authors to propose that the AT domains move independently during assembly line evolution, whereas the β-processing domains, ACP, and downstream KS domain evolve as a unit (**Figure 1.25**). This observation prompts a redefinition of the module at least in terms of PKS evolution: AT-(DH-ER-KR)-ACP_n-KS_{n+1} [124]. The same evolutionary comigration is also observed for the *trans*-AT PKSs [58, 125, 126], despite their distinct evolutionary history from the *cis*-AT PKSs. In addition, the alternative module definition is supported by certain results from PKS engineering, based on the fact that linker regions both up- and downstream of the AT domain (referred to as KAL (for KS-AT interdomain linker) and PAL1, respectively) are productive splice points for functional exchange of AT domains [127] (for further details, see **Figure 1.26**).

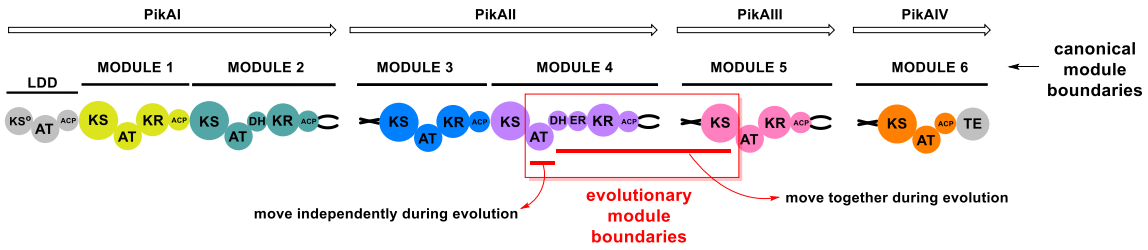


Figure 1.25 Canonical module definition as illustrated for the pikromycin PKS, and the revised definition of modules (red box) where putatively co-evolving domains (red underline) are highlighted. The canonical module definition matches the boundaries of unimodular subunits and corresponds to the functional unit for chain elongation. The evolution-inspired alternative module definition breaks the functional unit for chain elongation, but preserves the chain translocation unit (ACP_n/KS_{n+1}) [123].

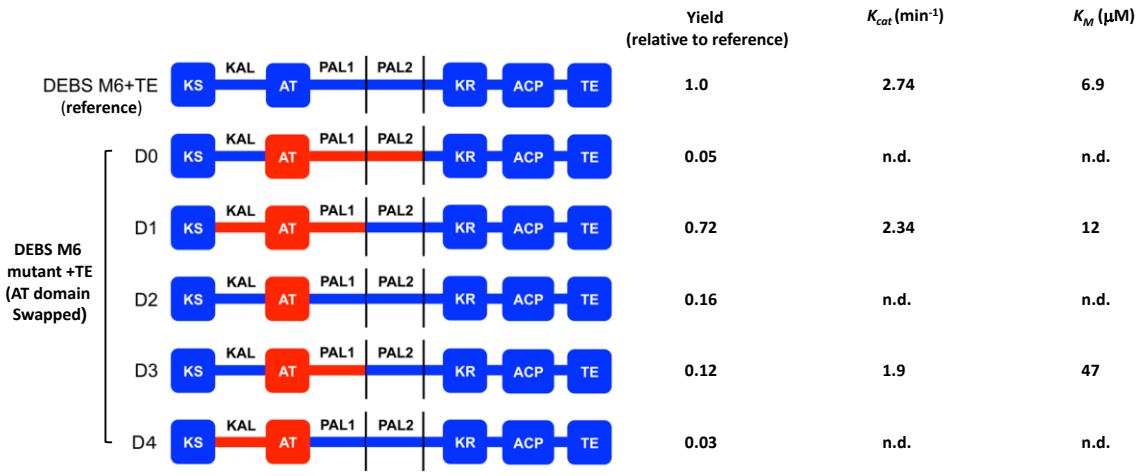


Figure 1.26 Scheme showing the strategy for AT exchange conducted by Yuzawa and co-workers [127]**.** The relative ability of DEBS M6+TE and the relevant mutants to catalyze chain elongation was evaluated based on enzyme kinetics. Compared to the wildtype M6 reference, the mutant D1 which carries the AT domain from EPOS (epothilone polyketide synthase) M4 along with its cognate KAL and PAL1 linkers showed only a slightly decreased k_{cat}, as well as a slightly increased K_M towards the native substrate. This observation supports the idea that the KAL and PAL1 region can be used as effective splice points for functional AT domain exchange.

Potentially transformative information as to useful splice sites was provided recently by a study reporting laboratory-based "Accelerated Evolution" [128]. In this experiment, the pKC1139-based toolkit was found to induce spontaneous genome rearrangement within several *Streptomyces* strains when the cultivation temperature was altered, an event which mimics evolutionary pressure. The result was the generation of a library of variants of the rapamycin PKSs from which modules had been unexpectedly deleted, inserted, and recombined (**Figure 1.27a**). Most surprisingly, the majority of these chimeric PKSs were active, and produced rapamycin analogues with titers comparable to those of wildtype. Detailed sequence analysis of the recombinants revealed several 'hotspots', most of which were located within conserved regions of the KS and AT domains, but additionally in the linker upstream of the ACP domains (**Figure 1.27b**). Whether the use of analogous sites can systematically improve rational PKS engineering in other systems, however, remains to be demonstrated.

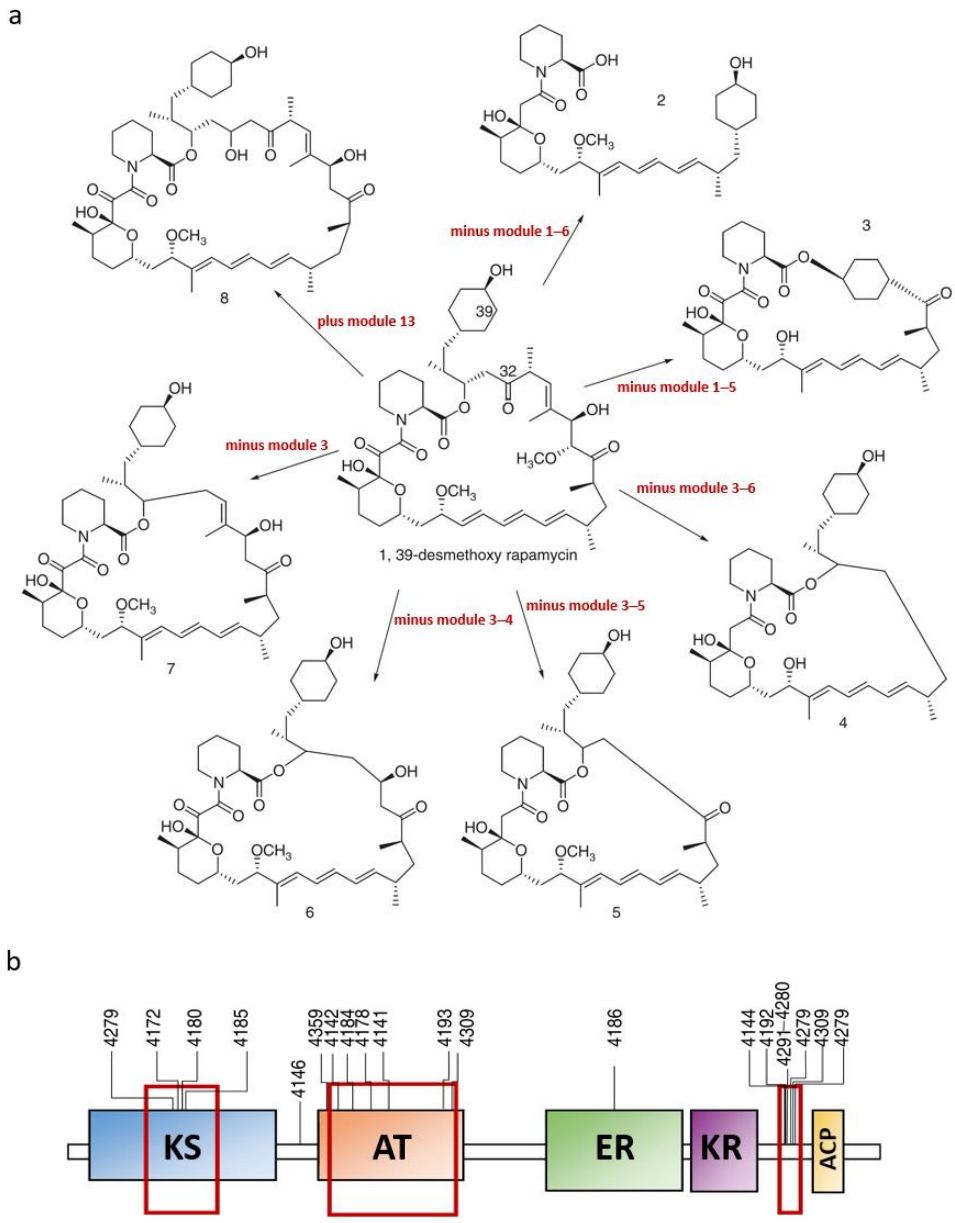


Figure 1.27 Rapamycin derivatives obtained via "Accelerated Evolution". a) Accelerated evolution of the rapamycin biosynthetic gene cluster led to recombination events resulting in net excision of certain modules or the addition of a second copy of module 13. **b**) Sequence analysis of the resulting constructs revealed potential recombination hotspots (shown in red boxes). These recombination junctions are generally located within the KS and AT domains or in the inter-domain linker region upstream of ACP domains (strain numbers are shown above the junctions). Adapted from reference [128].

These studies demonstrate the power of evolutionary-inspired engineering for producing functional chimeric PKSs, opening more possibilities of the rational design. However, we remain unclear of the overall dynamics of the comigration events and the

evolution direction between *cis*-AT and *trans*-AT. In terms of the proposed new definition of AT-(DH-ER-KR)-ACP_n-KS_{n+1}, it mostly corresponds to the evolutionary unit boundaries but conflicts to the functional and genetical boundaries. Does the new definition of module developed based on a few related systems is suitable for all *cis*-AT PKSs? Extensive evaluation is warranted before the new definition been accepted universally. For this, we engineered the stambomycin PKS based on both the classical and alternative module definition, which is described in **chapter II**.

1.7.5 Modifying post-PKS tailoring enzymes

In polyketide biosynthesis, post-PKS modifications contribute to the structural variation and biological activity of the final products. For the reasons mentioned earlier, tailoring enzymes, which typically act at the later steps in a pathway, are good candidates for synthetic biology. Functionally, post-PKS enzymes are classified according to the chemistry they catalyze (e.g. as transferases, oxygenases, halogenases, etc.). The following discussion highlights the structural diversification introduced by these types of enzymes.

Transferase enzymes include alkyl (usually methyl) transferases, aminotransferases, acyl (usually acetyl) transferases and glycosyltransferases, with both methyl- and glycosyltransferase being most common among post-PKS enzymes. The importance of this type of transformation is illustrated by the macrolide family of polyketides which often possess appended sugar moieties. These groups not only contribute to diversifying the structures, but also their biological properties, as they often interact directly with the cellular target, for example ribosomal subunits [129]. Work has shown that both the type and the number of sugars attached to the aglycone moiety significantly affect the bioactivity [130-134], and therefore glycosylation has often been targeted during attempts at post-PKS engineering. The majority of the sugar moieties are unique deoxy-hexoses which are derived from the key intermediates of sugar metabolism, such as NDP-4-keto-6-deoxyglucose. Normally, the genes involved in the activation and modification of the donor sugar are clustered together, an organization termed a sugar biosynthesis gene cassette, and present within the specialized metabolite gene clusters.

Generation of novel derivatives of glycosylated macrolides can be achieved by altering the sugar biosynthesis gene cassette or the substrate specificity of glycosyltransferases

(GTs), using mutagenesis or swapping [135]. In one example, hybrid macrolides were generated by combining the biosynthetic machinery for the macrolide tylosin which is decorated with three sugar moieties (D-mycaminose, L-mycarose and D-mycinose), with that of another macrolide, narbomycin, which incorporates D-desosamine. By expressing two genes (*nbmK* and *nbmJ*) involved in biosynthesizing D-desosamine in the tylosin producing strain *S. fradiae*, two novel tylosin analogues were obtained in which D-desosamine replaced D-mycaminose (**Figure 1.28**) [136]. In other work, by introduction of the GT gene *ole*G2 from the oleandomycin pathway into a mutant strain *Sac. erythraea* in which the *eryBV* (GT for mycarose sugar moiety), *eryCIII* (GT for desosamine sugar moiety) and *eryF* (CYP450) had been deleted, a new compound 3-*O*-rhamnosyl-6-deoxyerythronolide B was generated (**Figure 1.29**) [137].

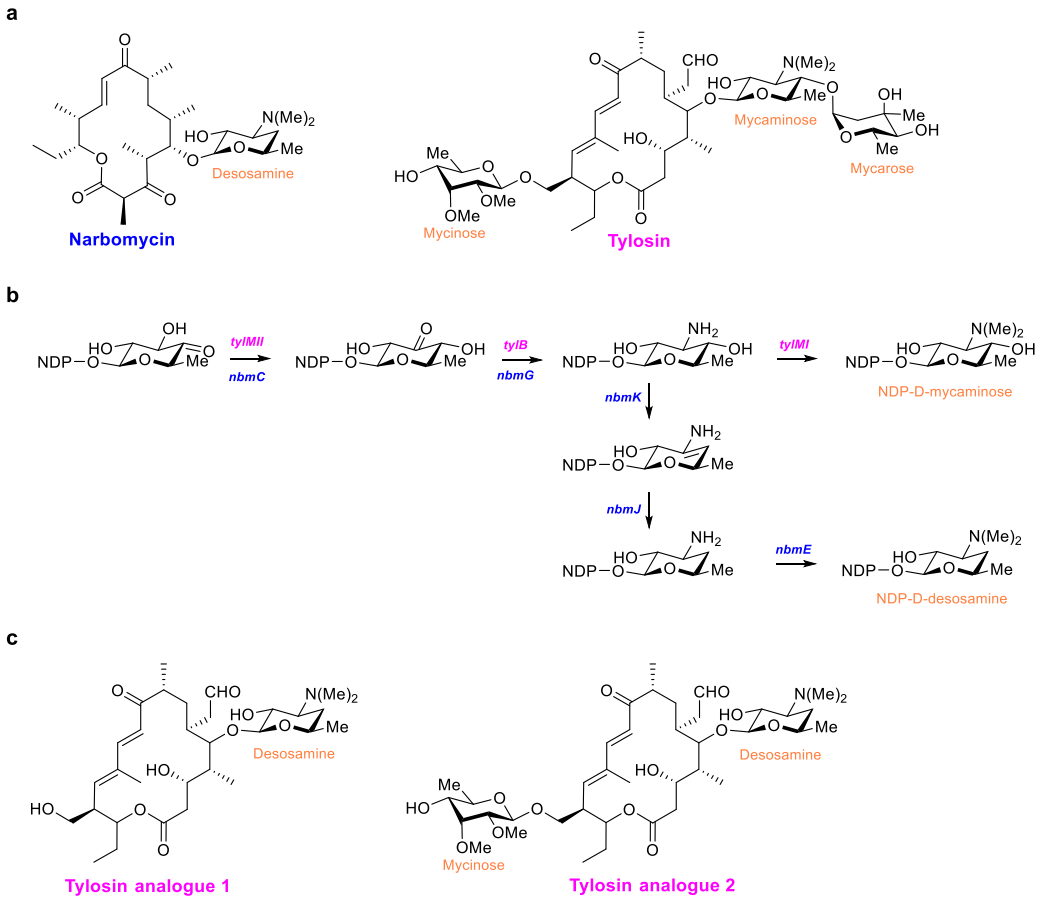


Figure 1.28 Alteration of a sugar moiety by modification of a sugar biosynthetic gene cassette. a) The structures of the native macrolide polyketides narbomycin and tylosin with their sugar moieties highlighted. **b**) Pathways for production of the sugars mycaminose and desosamine by the genes included in the BGCs

for tylosin and narbomycin, respectively. **c**) Two tylosin analogues produced by a mutant strain heterologously expressing genes *nbmK* and *nbmJ*. Adapted from reference [136].

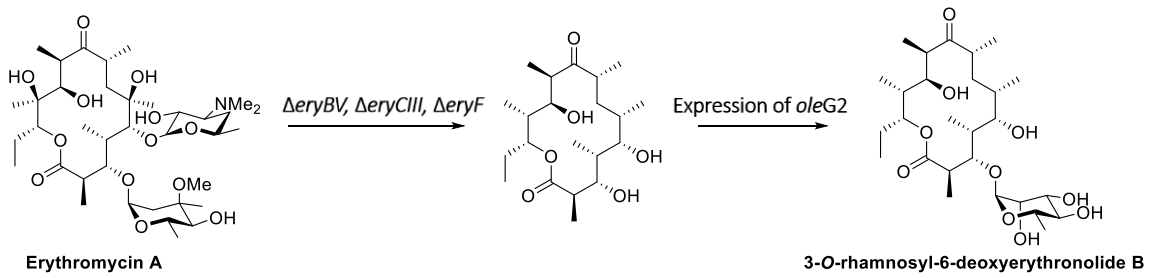


Figure 1.29 Alteration of an erythromycin A sugar moiety by complementation with alternative glycosyl transferases (GTs). Adapted from reference [137].

Oxygenases are also commonly observed post-PKS enzymes, which can catalyze varied chemistry including hydroxylation, epoxidation and oxidative rearrangement [135]. The most common among these enzymes are P450 monooxygenases, which catalyze regiospecific oxidation of the macrolide ring systems, providing an important layer of structural variability into the final natural products and often significantly influence biological activity [138, 139]. Due to their inherent substrate promiscuity, P450 monooxygenases had been used as a tool for generating various novel natural products. For example, the cytochrome P450 monooxygenase PikC from the pikromycin biosynthetic pathway in *S. venezuelae* can accept 12- and 14-membered macrolides as substrates, hydroxylating the macrolides at two different positions (**Figure 1.30a**). Indeed, when challenged with the product of a hybrid PKS in which the first Pik subunit PikAI was substituted by DEBS 1, the resulting analogues which differed from the normal PikC substrate at two ring positions were still accepted as substrates (**Figure 1.30b**) [140]. In some cases of rational diversification of polyketide structures, a P450-derived hydroxyl group can serve as an attachment point for subsequent glycosylation. This idea is exemplified by the generation of erythromycin analogues appended with a third sugar moiety. Specifically, macrolatones erythromycin A, C and D were fed to an *E. coli* strain in which the TDP-megosamine sugar biosynthetic pathway was overexpressed, giving rise to three analogues bearing megosamine at the C5-OH position (**Figure 1.31**). Relative to the parental macrolactones, these derivatives exhibited a marked increase in effectiveness against human and rodent malarial parasites (*Plasmodium falciparum* and *Plasmodium berghei*, respectively) [141].

Furthermore, the biochemical and structural studies on the enzymes themselves will also pave the way to a more rational approach employing genes encoding post-PKS tailoring enzymes. For instance, the structural comparison of several P450s particularly utilizing macrolides as substrates, such as CYP154C1 from *S. coelicolor* A3, CYP107W1 and CYP105D6 from *S. avermitilis*, PikC from *S. venezuelae*, and PimD from *S. natalensis* as well as MycG P450, has elucidated the conserved P450 folds characters and their respective characters towards substrate access pockets [142]. These study can help a better understanding of other P450 enzymes as well as their optimization and engineering for synthesizing novel polyketide derivatives.

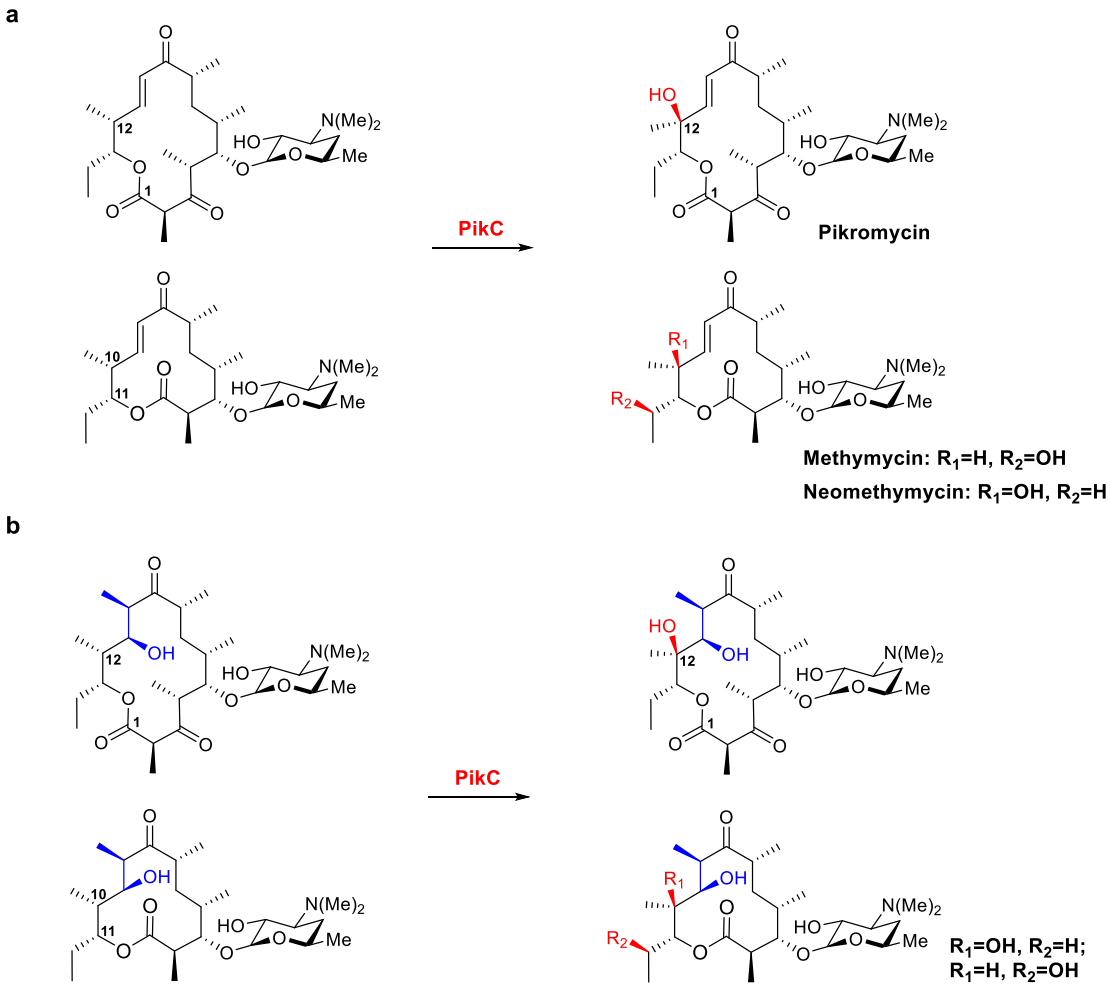


Figure 1.30 Generation of varied macrolactone structures using P450 monooxygenases. a) PikC-catalyzed hydroxylation on its native substrates in the biosynthetic pathway to pikromycin. b) PikC-catalyzed hydroxylation of novel substrate generated by a hybrid DEBS-Pik PKS pathway. Adapted from reference [140].

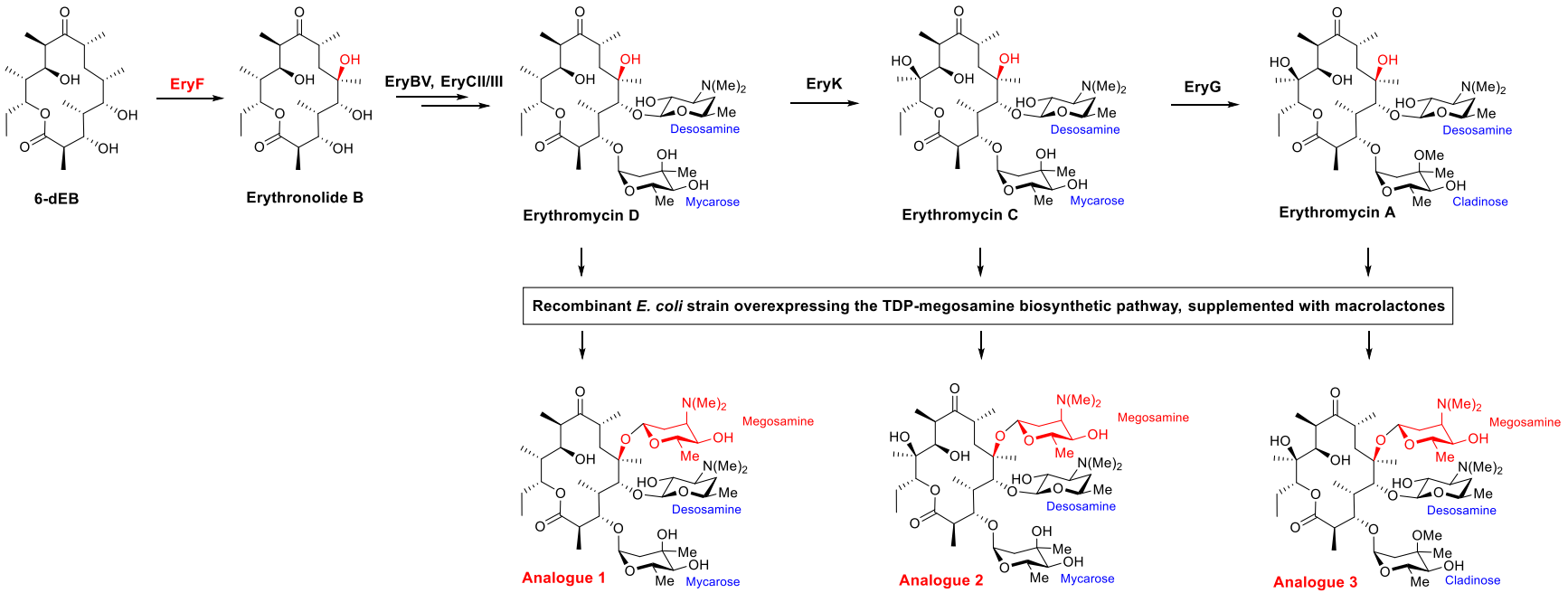


Figure 1.31 Generation of megosamine-containing erythromycin analogues by feeding the relevant erythromycin macrolactones to a strain of *E. coli* overexpressing the biosynthetic pathway to TDP-megosamine. Adapted from reference [141].

1.7.6 Altering monomer choice through precursor-directed biosynthesis and mutasynthesis

As described in **section 1.4.2.3**, precursor-directed biosynthesis (PDB) and mutasynthesis are promising approaches for further understanding and expanding PKS substrate promiscuity. In some cases, PDB and mutasynthesis rely on the intrinsic substrate tolerance of AT domains to accept fed analogues of their natural building blocks . In such cases, the fed molecules are CoA mimics such as synthetic SNAC thioesters, or free acids that are converted to CoA thioesters *in cellulo* by specific machinery such as MatB and CCR homologues [87, 143-145].

An illustrative example of using the PDB approach concerns the antimycins. Antimycins (ANTs) are a group of nine-membered NRPS/PKS hybrid products synthetized by *Streptomyces* sp. NRRL2288 and other *Streptomyces* strains (**Figure 1.32**). The natural ANTs differ in the alkylation at C7 and acylation at C8. The alkylation at C7 occurs via PKS-catalyzed elongation using variable alkylmalonyl-CoA extender units derived from the CCR-like protein AntE, while the acylation at C8 is introduced as a post-assembly line modification mediated by the promiscuous acyltransferase AntB. Mutasynthesis was carried out on this pathway by generating a Δ*antB* mutant which was deficient for ANT synthesis, but which accumulated dilactone C8-diacylated ANTs. When this strain was supplemented with a series of carboxylic acids, a panel of C8-diacylated ANTs derivatives was generated bearing various C7 substitutions, including halogen and alkyne termini, demonstrating that the precursors had been successfully activated in the cells by endogenous acyl-CoA ligases and β-oxidation, giving rise to the α,β-unsaturated CoA thioester substrates for AntE [145]. Similarly, Zhang and co-workers obtained another four C8-acylated ANTs analogues when the mutant Δ*antB*Δ*antE* was complemented with a mutated AntE (V350G) and fed with the corresponding acids (**Figure 1.32**) [146].

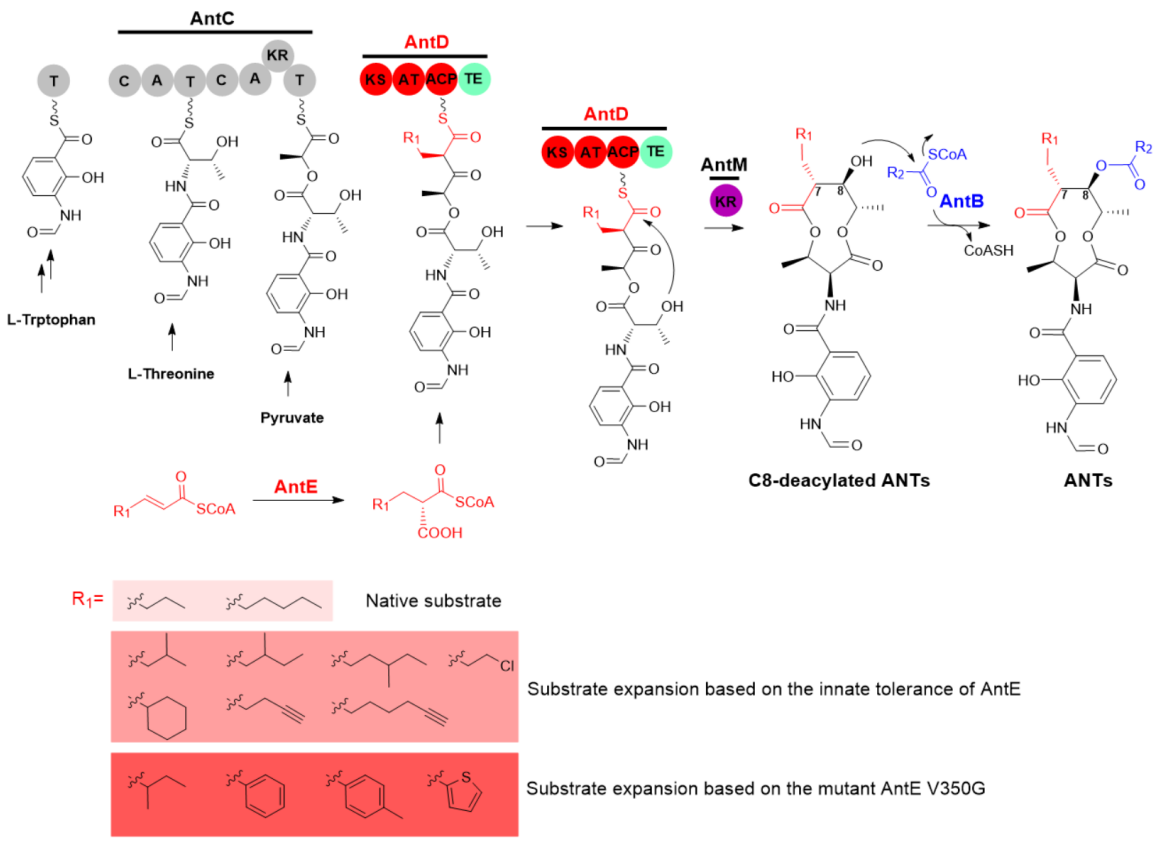


Figure 1.32 Biosynthetic pathway to the antimycins, and structural diversification based on the CCR type enzyme AntE. The NRPS/PKS hybrid assembly comprises subunits AntC and AntD. AntB catalyzes post-assembly line tailoring by acylation at C-8. AntM is a discrete ketoreductase. The origin of building blocks used to assemble the NRP/PK skeleton is shown. KS, ketosynthase; AT, acyl transferase; TE, thioesterase; C, condensation; A, adenylation; T, thiolation domain; KR, ketoreductase. Adapted from reference [145, 146].

Another notable mutasynthesis example concerns the structural engineering of FK506 and rapamycin. FK506 and rapamycin, which are both assembled by hybrid PKS/NRPS systems, are well known for their immunosuppressant and antifungal bioactivities. During FK506 biosynthesis, four special building monomers are incorporated into the skeleton: 4,5-dihydroxycyclohex-1-enecarboxylic acid (DHCHC), allylmalonyl-CoA, methoxymalonyl-ACP, and L-pipecolate, whose biosynthesis are all pathway-specific (**Figure 1.33**). With the exception of L-pipecolate that functions as an extender unit in the NRPS portion of the pathway, the other three extender units are incorporated by PKS machinery, with the DHCHC serving as a starter unit to prime the/ biosynthesis. The most unique structural feature of FK506 is the C-21 alkyl position, which derives from

allylmalonyl-CoA recruited by AT_4. Biosynthesis of this special monomer depends on a non-canonical pathway (shown in **Figure 1.11b**), involving four enzymes TcsA, B, C, D. In-frame deletion of the *tcsB* gene blocks biosynthesis of allylmalonyl-CoA, abolishing the production of FK506. Three genes, *anlE*, *F*, *G*, encoding enzymes responsible for isobutyrylmalonly-CoA biosynthesis in the biosynthetic pathway of ansalactam A, were then cloned and expressed in the Δ*tcsB* mutant. Due to the relaxed specificity of TcsD, the intermediate isobutyrylmalonyl-ACP was generated by AnlE, F, G from isobutyl-CoA, and was further dehydrogenated to yield isobutenylmalonyl-CoA (**Figure 1.33** panel **b**), which was subsequently installed at the C-21 position by AT_4, resulting in the final FK506 analogue [147]. Similarly, in-frame deletion of *fkbO* or *rapK* disabled the production of the starter unit DHCHC preventing synthesis of FK506 and rapamycin, respectively. The mutants were then utilized for mutasynthesis using synthetic analogues of DHCHC, resulting in numerous derivatives of FK506 and rapamycin carrying various cyclohexane moieties [148-151] (**Figure 1.33** panel **a**, exemplified by FK506).

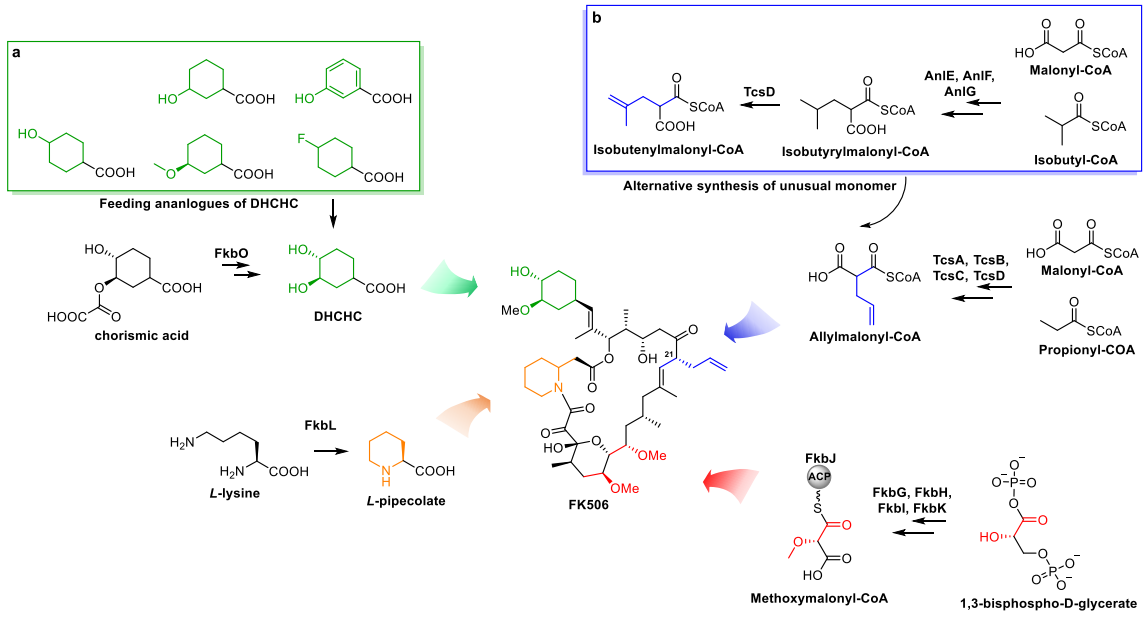


Figure 1.33 Pathways to monomers in the biosynthesis of FK506, and the relevant monomer alteration strategies for generating FK506 analogues. DHCHC, the PKS starter unit, is synthesized by FkbO using chorismic acid as substrate. Analogues of DHCHC, green frame in panel **a**, were used to feed a Δ*fkbO* mutant, resulting in analogues of FK506 containing various cyclohexane moieties. Allylmalonyl-CoA and methoxymalonyl-FkbJ, the PKS extender units, are generated by two sets of enzymes encoded by sub-gene clusters (the detailed conversion pathways are shown in **Figures 1.11b** and **1.12a**). Incorporation of three

enzymes from the ansalactam A biosynthetic pathway into the mutant Δ*tcsB* resulted in synthesis of the novel extender unit isobutenylmalonyl-CoA (shown in panel **b**), whose incorporation yielded FK506 derivatives at the C-21 position. L-pipecolate, the NRPS assembly line extender unit, is generated by FkbL using L-lysine as substrate. Adapted from reference [152].

Recently, a mutant of the MatB homologue from *Rhizobium trifolii* was demonstrated to be able to convert a range of malonate derivatives to their relevant CoAs, including allylmalonyl-CoA, propargylmalonyl-CoA, and azideoethyl-CoA [153]. Subsequently, Musiol-Kroll and co-workers expressed an engineered mutant of MatB (MatB T207G/M306I) in *S. collinus* Tü 365, a kirromycin producing strain, and obtained the allyl- and propargyl-incorporated kirromycins by adding exogenous allylmalonic acid and propargylmalonic acid to the culture broth (**Figure 1.34a**) [154]. The latter extender unit was utilized as a target for further modification by 'Click Chemistry'. More recently, a MatB-type enzyme was discovered from the monensin producer *S. cinnamonesis* and shown to possess activity *in vivo* towards a small number of malonic acid derivatives, generating malonyl-CoA derivatives which were incorporated more efficiently than the corresponding SNAC derivatives [155]. Subsequently, the allyl-, propyl-, and propargyl-monensins were successfully produced when the corresponding malonate derivatives were added into the fermentation culture of *S. cinnamonensis* ATCC15413, accompanied by co-expression of MatB (**Figure 1.34b**) [144].

Given the remarkable substrate promiscuity of CCRs and MatB-type enzymes and their mutants, these enzymes hold promise as a toolkit for generating a wide range of unnatural building blocks and their downstream polyketide derivatives, via PDB and mutasynthesis approaches. Notably, new analogues incorporating alkyne functionality are attractive targets for further structural derivatization by "Click Chemistry", which can facilitate subsequent investigations into modes-of-action [156].

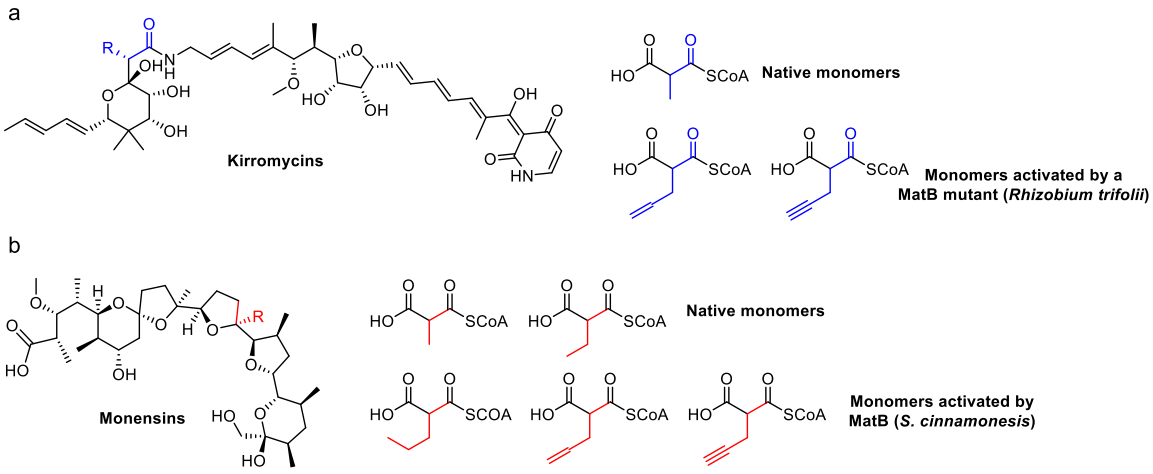


Figure 1.34 Structures of monensin and kirromycin and their analogues generated by incorporating alternative monomers. Adapted from references [144, 153-155, 157].

1.8 Stambomycins and the stambomycin PKS

Streptomyces ambofaciens ATCC23877 was isolated in 1954 in Peronne, France [158], and was initially described to produce congocidine and spiramycin. Similar to other streptomycetes, *S. ambofaciens* ATCC23877 has a large linear chromosome (about 8Mbp). At the extremities, the chromosome presents 200Kb of terminal inverted repeated regions (TIR) [159].

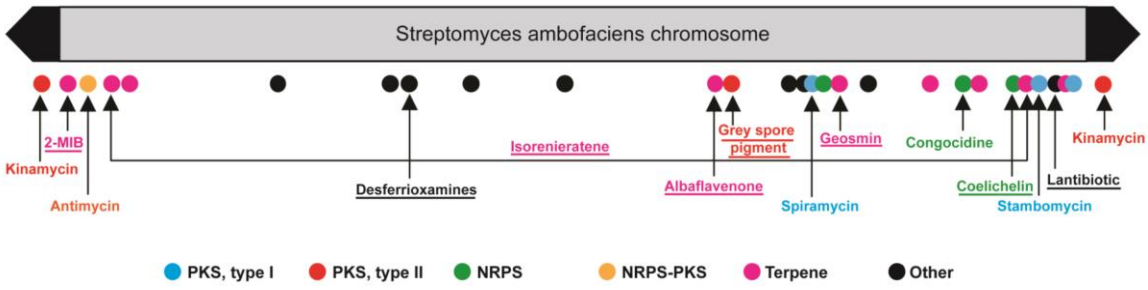


Figure 1.35 Overview of the linear chromosome of *S. ambofaciens* ATCC23877 and the locations of BGCs identified to date. Clusters are indicated by filled circles, whose colors correspond to the type of genes found in the cluster. Metabolites in S. ambofaciens which are identical to S. coelicolor are underlined. The black arrows at the extremities of the chromosome arms symbolize the terminal inverted repeats (TIRS).

Mining the genome of this strain revealed roughly 30 BGCs for secondary metabolites (**Figure 1.35**), in which the stambomycins (A–F) are a group of macrolide polyketides been identified in this approach [160, 161]. Initially, *in silico* analysis revealed a large type I PKS gene cluster located in the right arm of the chromosome of *S. ambofaciens*, but no

corresponding product was observed, implying that the type I PKS gene cluster was silent or poorly expressed under standard culture conditions. Therefore, a method to specifically activate expression of this cryptic cluster was developed in order to access the potential novel macrolide. Based on DNA sequence analysis, three genes encoding putative transcriptional regulators were identified within the type I PKS gene cluster: *samR0468* (a response regulator), *samR0469* (a histidine kinase), and *samR0484* (a member of the LAL (Large ATP-binding LuxR) family proteins). As members of LAL family regulators had already been reported as activators of macrolides biosynthesis, including for example PikD of *Streptomyces venezuelae* and RapH in *Streptomyces hygroscopicus* which activate respectively pikromycin and rapamycin production, the gene *samR0484* was cloned into the integrative vector pIB139 and overexpressed under control of the constitutive promoter *ermEp**. Following this manipulation, the strain containing the LAL overexpression plasmid produced the expected metabolite, a 51-membered macrolide polyketide, while the control strain containing empty plasmid pIB139 did not.

Subsequently, sequence alignment showed that genes *samR0464* and *samR0488* encode respectively an endoribonuclease and a "cold-shock" DNA-binding protein, which are unrelated to the biosynthesis of stambomycins confirmed through gene deletion, thus indicating the putative boundaries of the PKS gene cluster. Between *samR0487* and *samR0465*, the gene cluster contains 25 genes and spans almost 150 kb (**Figure 1.36a**), including 9 genes encoding PKS multienzymes (which together comprise 25 modules, themselves composed of 112 enzymatic domains) and 16 additional genes that encode proteins predicted to be involved in PKS substrate supply, post and on-PKS tailoring reactions, sugar moiety biosynthesis, regulation and resistance. With their 51-membered macrolactone ring core and gigantic biosynthetic gene cluster, the stambomycins are among the largest and therefore most complex polyketides identified to date.

Among these additional genes, *samR0487*, *0486*, *0473*, *0480* and *0472* were found to be homologous to genes *srm16, 17, 14, 15* and *21* in the spiramycin BGC that are responsible for the biosynthesis of TDP-D-mycaminose. In addition, the *samR0481* gene was shown to encode a glycosyl transferase (GT) capable of transferring the mycaminosyl residue from NDP to hydroxyl groups of the stambomycin macrolide [162]. The two other biosynthetic genes encoding cytochrome P450 enzymes (*samR0478* and *samR0479*) were

confirmed to catalyze hydroxylation reactions, with SamR0479 being responsible for C-terminal hydroxylation required for offloading of the fully assembled polyketide chain from the stambomycin PKS via macrolactonization, while SamR078 introduces the hydroxyl at C-28 [163]. This novel mechanism for macrolactone formation based on a P450-installed hydroxyl group, rather than that generated by ketoreductase, represents a second unique characteristic of the stambomycin biosynthetic pathway. The gene cluster also contains *samR0475* encoding a putative type II thioesterase, which is presumed to have a proof-reading function via release of aberrant acyl groups attached to the ACP domains during the chain elongation process [164, 165].

Another important characteristic of the stambomycins is the diversity of side chains present at C-26, which derive from the incorporation of six atypical 4-methylhexyl-/5-methylhexyl-/4-methylpentyl-/hexyl-/pentyl-/4,5-dimethylhexylmalonyl-CoA by the unusual AT domain on module 12 (**Figure 1.36a** red box). Recently, Ray and co-workers demonstrated that SamR0482 (acyl-CoA synthetase) and SamR0483 (acyl-CoA carboxylase) together are involved in the biosynthesis of these unusual medium chain acyl-CoAs (**Figure 1.36c**, the pathway was described in detail in **section 1.6.2.1**).

Bioactivity tests indicated that the stambomycins present promising anti-tumor activities against multiple cancer cell lines, with measured IC_{50} values similar to or even superior to those of several drugs in clinical usage. For example, the metabolites inhibited proliferation of human adenocarcinoma (HT29) cell lines with similar IC_{50} values (1.8 ± 0.04 μM for stambomycins A/B and 1.7 ± 0.04 μM for stambomycins C/D) to the clinical anticancer agent doxorubicin (1.3 ± 0.08 μM), but were 4-fold less toxic toward adult Chinese hamster ovary cells (CHO-K1; 8.5 ± 0.7 μM for stambomycins A/B, 8.5 ± 0.5 μM for stambomycins C/D) than doxorubicin (2.0 ± 0.2 μM) [160]. These data highlight the stambomycins as promising new leads for anticancer drug discovery.

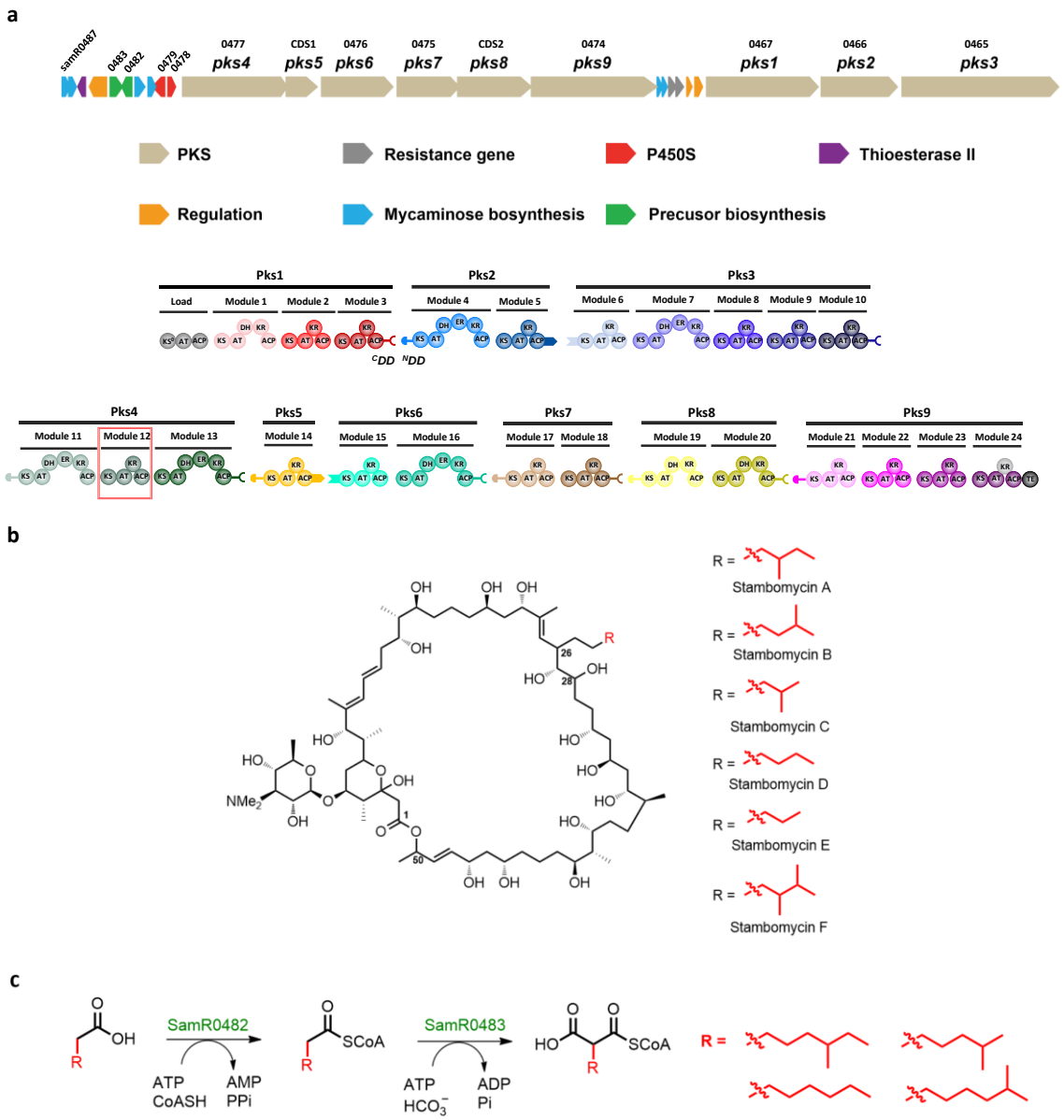


Figure 1.36 Biosynthesis of the stambomycins in *S. ambofaciens* ATCC23877. a) Organization of the stambomycin biosynthetic gene cluster showing the domains, modules, subunits and docking domains. The AT domain of PKS module 12 (red boxed) recruits six alternative extender units, resulting in the small family of stambomycins. The last KR domain on module 24 (grey) is inactive. **b**) Structures of the stambomycins A–F, showing the differences in their side chains at C-26. The configurations of the α- and β -stereocenters generated by each KR domain was predicted based on Keatinge-Clay's model [166]. **c**) The elucidated biosynthetic pathway to the unusual stambomycin PKS extender units. SamR0482, an acyl-CoA synthetase, catalyzes the activation of precursor originating from fatty acid metabolism using the cofactor ATP and coenzyme A. SamR0483, an acyl-CoA carboxylase, catalyzes the subsequent ATP-dependent carboxylation reaction, resulting in the final extender units. Adapted from reference [85, 160, 161].

1.9 Thesis objectives

This thesis was funded by the IMPACT Biomolecules project of the I-site ("*Initiatives – Science – Innovation – Territoires – Economie*") Lorraine Université d'Excellence (LUE). Globally, the project aims to identify and generate new biomolecules with anti-inflammatory, anti-proliferative and antioxidant properties for diverse industrial applications. It involves different workpackages (WPs) ranging from natural product discovery (WP1), to screening bioactivities against specific targets and functionalization to improve the potential of promising hits (WPs 2 and 3), to formulation/delivery studies (encapsulation and vectorization) to facilitate the transition into industrial application (WP2).

Work presented in the thesis relates to the transversal action 1 (TA1: polyketide), which focused on three main aspects: TA1-i) annotation and analysis of new gene clusters identified by WP1; TA1-ii) improvement of titers of these novel compounds to permit synthetic biology, chemical derivatization/formulation (by WP2) and biological testing (by WP3); and TA1-iii) genetic engineering of polyketide biosynthetic pathways to produce specific derivatives for downstream biological evaluation (WP3) and further optimization/vectorization (WP2). In this context, this PhD project addressed the third aim, and was based on molecules in hand, for which the approaches developed and insights obtained could applied to additional polyketide systems identified by WP1.

The model system used to obtain this critical proof-of-principle for our approach was stambomycin PKS. Specifically, we chose to focus on two principal approaches: i) mutasynthesis; and ii) exchanging of intact PKS units (both modules and subunits) to alter the structure of the stambomycins in a predictable way. The overall thesis is divided into three results chapters II, III, IV. Chapters II and III describe the two principal projects anchored in PKS engineering, while chapter IV describes an additional project concerning desferrioxamine metabolites, which directly resulted from the other two.

Chapter II Generation of ring-contracted stambomycins by modular polyketide synthase engineering

The research work of this chapter has been submitted to Nature Communications with the title "Successes, surprises and pitfalls in modular polyketide synthase engineering: generation of ring-contracted stambomycins"(DOI:10.21203/rs.3.rs-222743/v1), and is currently under review.

2.1 Summary of this work

In modular type I PKSs, the direct relationship between domain organization and product structure forms the basis for prediction of polyketide architectures from their encoding genes, stimulating efforts to engineer PKSs to produce new polyketides in a predictable way. Despite decades of efforts, genetic engineering of polyketide synthase multienzymes remains a difficult task, as when the desired products are obtained their yields are reduced and often substantially relative to the parental metabolites, or the experiments simply fail for no evident reason (see **Chapter I**, **section 1.7**). Recently, two structural studies on PKS modules help to support the idea that the most productive strategy for further engineering efforts will be to employ modules/subunits as the basic building blocks for engineering, due to the high functional interdependence of the component domains. To do so, both covalent linkers and non-covalent docking domains could in principle be used to generate novel combinations of such units [112, 167].

Docking domains (DDs) are short regions at the extreme ends of the PKS subunits which have been shown previously to ensure the correct subunit-to-subunit interactions leading to the desired polyketide product [55], and which, critically, also exhibit a certain portability into alternative contexts [63]. DD engineering has been used successfully to both artificially split and splice PKSs (see **Chapter I**, **section 1.7**), but most of the experiments have been carried out *in vitro*, with little research targeting full PKS assembly lines. Thus, our efforts focused on developing strategies to exchange large sections of PKS assembly lines, and in particular whole subunits, by engineering of docking domain-based communication using stambomycin modular PKS as a model system. This gigantic PKS assembly line consists of 9 PKS subunits and therefore 8 docking domain pairs. Indeed, given the extraordinary size of the PKS, the stambomycins, at 51 members, are among the

largest macrocyclic polyketides identified to date. These features make the stambomycin PKS an ideal system to investigate and exploit the basis for interprotein communication, and the products an attractive model for structural simplification to yield a series of smaller, cyclic derivatives (potential 'minimal pharmacophores') for biological evaluation.

Specifically, we engineered a new interface between subunits Pks4 and Pks9 (**Fig. 1** of the submitted manuscript) using an approach combining modification of docking domains and a set of key interdomain interactions underlying chain transfer and extension. The choice of this specific alternative junction was based on three considerations: i) analysis of DDs showing that those operating at the subunit 4/5 and subunit 8/9 inter-PKS junctions belong to the same structural class, and thus there should be an intrinsic compatibility; ii) the incoming substrate for the first catalytic KS domain of subunit 9 (KS_{21}) and that of subunit 5 (KS_{14}) are similar over the first several carbons, and thus we anticipated that KS_{21} would tolerate the non-cognate substrate provided by subunit 4; and, iii) Pks4 carries the AT_{12} domain responsible for the structural variation among the stambomycin family members, and thus keeping Pks4 would result in the corresponding ring-contracted stambomycins, providing important initial validation for the success of the engineering.

We initially created the new junction between Pks4 and Pks9 both by wholesale swapping of helix α3 on $^{C}DD_4$ for that of $^{C}DD_8$ resulting in a $^{C}DD_4$ helix swap, and by mutating specific residues in $^{C}DD_4$ which were considered to be likely specificity determinants to match those of $^{C}DD_8$ (referred to as $^{C}DD_4$ SDM, SDM for site-directed mutagenesis) therefore rendering it compatible, at least in principle, with $^{N}DD_9$. These modifications were made both in the context of a deletion of the intervening subunits (Pks5−8) to remove competition for binding of Pks4 by Pks5 and of Pks9 by Pks8, and with them present.

HPLC-MS analysis of the fermentation extracts of these generated mutants indicated that stambomycins were still produced from mutants in which the intervening subunits (5−8) were preserved. In contrast, mutants in which subunits 5−8 were removed lost the ability to produce stambomycins, as expected. None of the expected ring-contracted stambomycins were detected in these initial experiments. Instead, a series of shunts products were obtained, which corresponded to polyketide intermediates released from

modules 12 and 13 (**Fig. 1** of the submitted manuscript). This observation reflects the fact that newly created junctions are likely nonfunctional. Further *in vitro* investigation of DD pairs by isothermal titration calorimetry (ITC) revealed that 'SDM' modification was sufficient to disrupt native docking but inadequate to induce a new interaction (i.e. $^{C}DD_4$ SDM was not able to dock with either $^{N}DD_5$ or $^{N}DD_9$), while the '$^{C}DD_4$ helix swap' strategy resulted in a new, productive docking interface ($^{C}DD_4$ helix swap notably has closely similar docking affinity to $^{C}DD_8$ for $^{N}DD_9$). Nevertheless, the new interface 'ACP_{13}-$^{C}DD_4$ helix swap/$^{N}DD_9$-KS_{21}' failed to function *in vivo*. Taken together, these observations reinforced the notion that the ACP_n/KS_{n+1} domain pair needs to be taken into consideration during interface engineering [111], as this protein-protein interaction contributes to the overall chain transfer reaction.

In this context, we attempted to optimize the engineering strategy by improving the compatibility of ACP_{13}/KS_{21} as well as by maintaining the native ACP_{13}/KS_{14} junction to facilitate substrate translocation. For this: i) helix α1 of ACP_{13} was modified to match that of ACP_{20} (the native partner of KS_{21}); and, ii) both $^{N}DD_9$ and the first domain (KS_{21}) of subunit 9 were replaced by that of $^{N}DD_5$+KS_{14} using two different splice sites within the KS domain (the site located at the end of KS domain represents a conserved motif that was exploited recently for efficient swapping of the downstream AT domain, while the other which is located near the middle of the KS domain, is a splicing hot spot identified by induced recombination within several PKS systems [127, 128]. This strategy resulted in three modified interfaces between subunits 4 and 9: 'ACP_{13} SDM-$^{C}DD_4$ helix swap/$^{N}DD_9$-KS_{21}', 'ACP_{13}-$^{C}DD_4$/$^{N}DD_5$-KS_{14}' and 'ACP_{13}-$^{C}DD_4$/$^{N}DD_5$-$KS_{14/21}$'. In this way, we successfully obtained the target 37-membered stambomycins (**Fig. 5** of the submitted manuscript).

The yields of ring-contracted stambomycins were dramatically reduced compared to the native 51-membered stambomycins, and instead, the modified systems produced substantial quantities of shunt products corresponding to intermediates generated by the PKS up to Pks4. Interestingly, we could demonstrate by targeted inactivation, that the C-terminal TEI domain of Pks9 plays an active role in release of these chains, representing a rare example of intersubunit cross-talk. In principle, the reason for stalling of intermediates

upstream of the newly-generated Pks4/9 interface could be the limited rate of chain extension in the first module of Pks9, as in this case, there is potential incompatibility between the swapped KS_{14} and hybrid $KS_{14/21}$ and the partner ACP_{21}. Subsequently, therefore, we introduced modifications to ACP_{21} to improve its compatibility with the upstream KS domain, but no major improvement in yields was obtained.

Overall, we have carried out the most dramatic modification to an intact PKS assembly line reported to date, removing a total of seven internal modules. Nonetheless, the low yields of products from the modified assembly line demonstrate that even engineering strategies which have been 'optimized' based on precedent literature do not overcome the many intrinsic barriers to efficient functioning of modified assembly lines. In particular, our work highlights that the poor substrate tolerance of modules downstream of introduced junctions can represent a significant block to successful PKS synthetic biology. Future work in the area should therefore address ways to boost their tolerance (or alternatively identify intrinsically broad specificity alternatives), as well as the molecular basis for the TEI-mediated cross-subunit off-loading, as this mechanism, if general, will also further reduce yields of target metabolites.

2.2 Results

Manuscript of the submitted work

Successes, surprises and pitfalls in modular polyketide synthase engineering: generation of ring-contracted stambomycins

Li Su[1,2], Laurence Hôtel[2], Cédric Paris[3], Alexander O. Brachmann[4], Jörn Piel[4], Christophe Jacob[1]*, Bertrand Aigle[2]*, Kira J. Weissman[1]*

[1]Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France. [2]Université de Lorraine, INRAE, DynAMic, F-54000 Nancy, France. [3]Université de Lorraine, PASM, F-54000 Nancy, France. [4]Institute of Microbiology, ETH Zurich, 8093 Zurich, Switzerland.

Correspondence and requests for materials should be addressed to C.J. (e-mail: christophe.jacob@univ-lorraine.fr), B.A. (e-mail: bertrand.aigle@univ-lorraine.fr) or K.J.W. (e-mail: kira.weissman@univ-lorraine.fr)

Abstract

The modular organization of the type I polyketide synthases (PKSs) would seem propitious for rational engineering of desirable analogous. However, despite decades of efforts, such experiments remain largely inefficient. Here, we combined multiple, state-of-the-art approaches including modification of docking domains, use of modules of varying domain composition, alternative interdomain fusion sites, and targeted adaptation of key domain-domain interfaces, to reprogram the stambomycin PKS by deleting seven internal modules – the most substantial modification to an intact system reported to date. One such system produced the target 37-membered mini-stambomycin metabolites, a reduction in chain length of 14 carbons relative to the 51-membered parental compounds, but also substantial quantities of shunt metabolites released from the multienzyme subunit upstream of the newly-installed junction. Our data also provide evidence for an unprecedented off-loading mechanism of such stalled intermediates involving the C-terminal thioesterase domain acting on chains located four modules upstream. The yields of all metabolites were substantially reduced compared to the wild type compounds, likely reflecting the poor tolerance to the non-native substrates of the modules downstream of the introduced interfaces. Taken together, our data demonstrate that even 'optimized' PKS engineering strategies remain inadequate for efficient production of target polyketide derivatives, and highlight several areas for future investigation.

Introduction

For almost thirty years, efforts have been made to leverage the modular genetic architecture of the type I polyketide synthases (PKSs) to generate novel derivatives, typically by modifying individual catalytic domains. Despite enormous progress in establishing domain structure-function relationships[1,2], such genetic manipulation remains inefficient[3]. Insight into factors potentially contributing to low product yields was provided by cryo-electron microscopy analysis of a model PKS module at multiple stages of its catalytic cycle[4,5]. This work revealed that interdomain contacts are critical for establishing the various functional

states of the module, and that transitions between such states rely on evolving interfaces between the domains, as well as the intervening 'linker' regions. In short, PKS modules appear to be highly integrated units, thus explaining why exchange of catalytic domains for heterologous counterparts is often detrimental[6]. Collectively, these observations motivate future approaches in which modules or multi-modular subunits are employed as the basic building blocks for engineering the assembly lines[7–13].

Nevertheless, using modules requires a clear definition of their domain composition. Traditionally, modules encompass the three invariable domains required for monomer selection and chain extension (ketosynthase (KS), acyl transferase (AT) and acyl carrier protein (ACP)), as well as any intervening β-keto processing activities (e.g. ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)) (**Fig. 1a**). However, an alternative definition was recently suggested by the finding that KS domains in certain PKSs co-evolve with the tailoring domains located upstream in the assembly lines[14,15]. Accordingly, modules begin with the modifying domains and the associated AT, and terminate with the KS that is classically assigned to the downstream module (**Fig. 2**). Even before a module redefinition was suggested, engineering efforts revealed that maintaining the key ACP_n/KS_{n+1} interface can, in certain cases, be critical for the function of a hybrid PKS[7,10,16]. Recently, we have carried out module swapping based on both of these definitions, by covalently tethering heterologous modules to a common donor module within a bimodular mini-PKS[17]. Overall, our data demonstrated that both module definitions led to functional hybrid PKSs, and which boundaries worked best depended on the source module[18]. Indeed, regardless of which extremities are employed, module exchange results in non-native interdomain interactions (ACP_n/KS_{n+1} or KS_{n+1}/ACP_{n+1}), and in the case of classical module boundaries, potential incompatibilities in terms of KS substrate specificity (**Fig. 2**) – both of which have been shown to reduce activity via detailed studies *in vitro*[19–21].

In this work, we aimed to investigate the generality of these findings for efforts to create non-native intermodular junctions when the modules are located on distinct subunits[10]. In such cases, the resulting non-covalent interactions are mediated by short sequences at the extreme C- and N-termini of the subunits called docking domains (DDs)[22] (**Figs. 1** and **2**). Matched pairs of DDs form specific complexes at intersubunit interfaces, enforcing a strict subunit ordering within the PKS system. As a test case, we aimed to genetically engineer the biosynthesis of substantially smaller derivatives of the stambomycin family of polyketides in *Streptomyces ambofaciens* ATCC23877[23]. The stambomycins **1** (**Fig. 1b**) are glycosylated macrolides which show promising anti-cancer activity against multiple human cancer cell lines[23]. The six characterized family members (A−F) differ in the alkyl functionality at the C-26 position (**Fig. 1a**) which directly impacts their potency,[23,24] but have in common modification by *trans*-acting cytochrome P450 hydroxylases at C-28 and C-50. Notably, at 51-members, the macrolactone ring is among the largest of all known polyketides. Thus, the stambomycins represent an attractive model system for establishing PKS engineering as a means to access structurally-simplified analogues (minimal pharmacophores[25]) for biological evaluation, as a complement to traditional chemical synthesis.

Here we report a comprehensive series of experiments aiming to generate 37-membered ring stambomycin analogues, based on both the classical and revised module definitions. The target mini-stambomycins were detected successfully, albeit in low yields, and the identification of the derivatives allowed us to clarify the relative timing of the two cytochrome P450-catalyzed hydroxylation reactions in the pathway. Attempts to boost titers by ACP/KS interface engineering[19,20,26] were unsuccessful, but led in certain cases to a surprising increase in the liberation of linear shunt metabolites – only the second report, to our knowledge, of inter-subunit crosstalk resulting in thioesterase (TE)-mediated chain release. Taken together with recent work by others[10], our data reinforce the idea that in order to boost efficacy, strategies based on modifying PKS intersubunit interfaces must take into account the function of the modules acting downstream from the newly-established junctions.

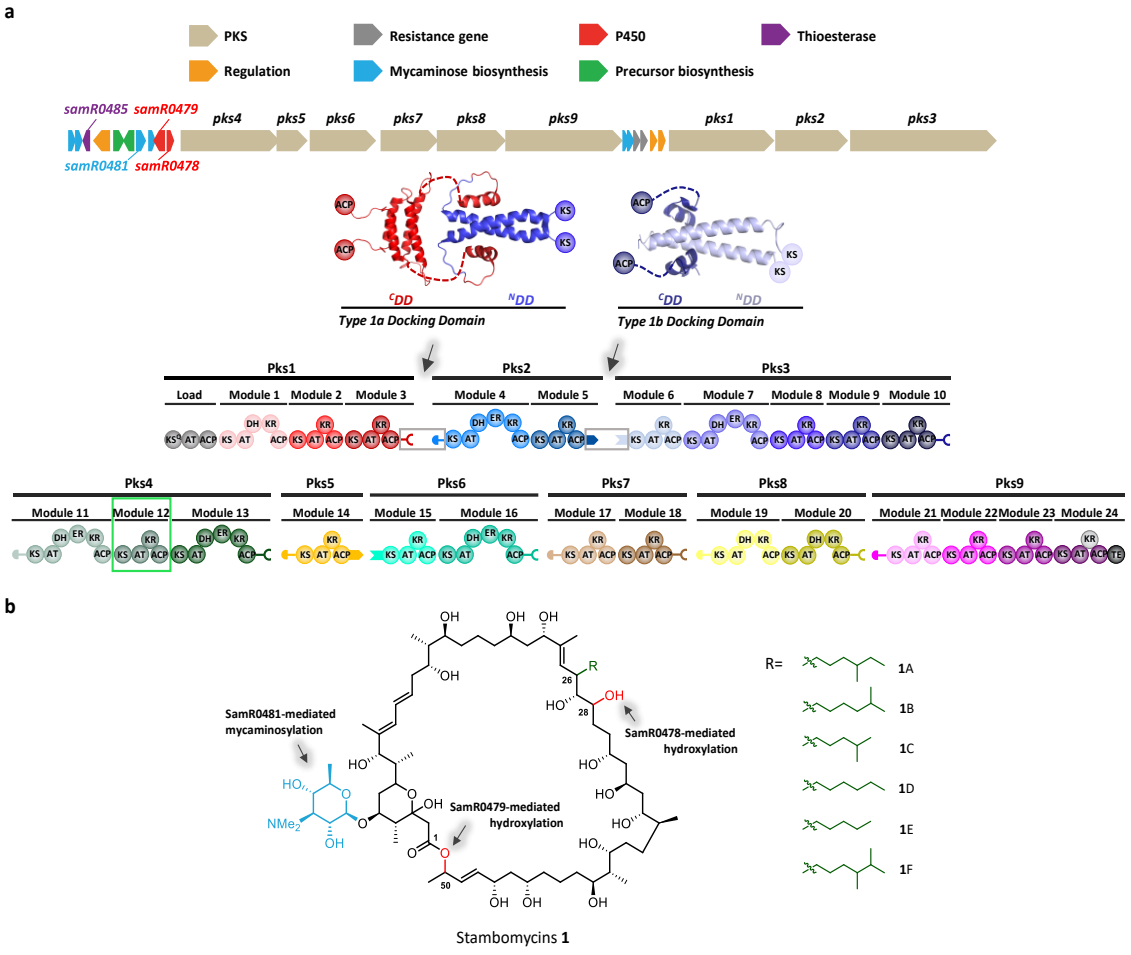


Fig. 1 Stambomycin polyketide synthase (PKS) and structures of stambomycin derivatives produced by *S. ambofaciens* ATCC23877. a Organization of the stambomycin biosynthetic gene cluster, and schematic of the encoded PKS subunits (Pks1–Pks9) showing the component modules and domains, as well as the intersubunit docking domains. The DDs belong to two distinct structural classes (type 1a and type 1b), for which representative NMR structures of complexes are shown[22,27]. The AT domain of PKS module 12 (green box) is responsible for recruiting six alternative extender units, resulting in a small family of stambomycins. The last KR domain of module 24 (grey) is inactive. **b** Structure of stambomycins **1** (A–F), which differ from each other in the alkyl functionality (R group) at position C-26

(the indicated stereochemistries[23] have been predicted based on analysis of known domain stereochemical determinants[66], but have not been directly confirmed). The sites of glycosylation and hydroxylation are highlighted with their responsible enzymes indicated. Abbreviations: AT, acyl transferase; KS, ketosynthase (KS^Q refers to replacement of the active site cysteine residue by glutamine); ACP, acyl carrier protein; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; TE, thioesterase; ^{C}DD, C-terminal docking domain; ^{N}DD, N-terminal docking domain.

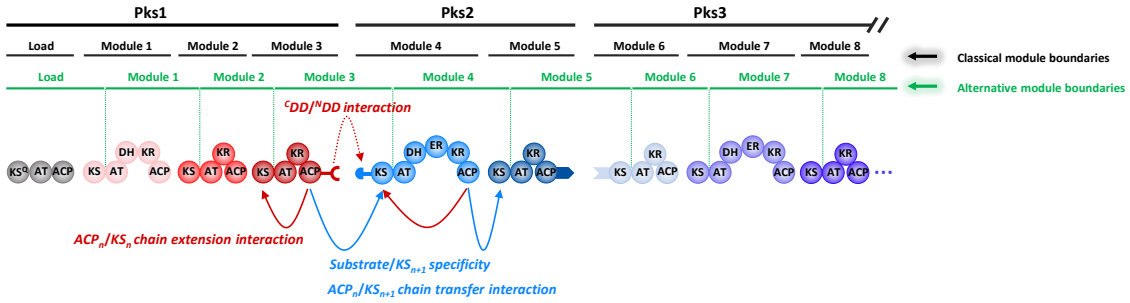


Fig. 2 A schematic representation of classical and alternative module definitions. A classical module (black) is defined as the catalytical unit responsible for incorporation of one building block into the growing polyketide chain, and associated functional group modifications. ^{C}DD/^{N}DD pairs (shown) mediate communication between such traditionally-defined modules. According to this definition, KS/ACP chain extension interactions (left-pointing arrows) occur within the modules, whereas ACP/KS chain transfer interactions (right-pointing arrows) occur between modules, and the incoming substrate for the KS domain is generated by the upstream module. The alternative module boundaries (shown in green) were inspired by the evolutionary co-migration in certain systems of the KSs and the upstream processing domains[14,15]. Under this revised definition, the KS/ACP chain extension interaction (left-pointing arrows) is inter-modular, while the ACP/KS chain transfer contacts (right-pointing arrows) are intra-modular.

Results

Design of engineering experiments based on classical modular boundaries. The stambomycin PKS comprises 25 modules distributed among 9 polypeptides (Pks1−9)[23] (**Fig. 1a**) (Note: throughout the text, the stambomycin genes have been numbered in accordance with ref. [23]). To access abridged derivatives using the classical module boundaries, we reasoned that we could engineer novel intersubunit interfaces by suitable manipulation of docking domains. Encouragingly, the extreme C- and N-termini of all subunits (with the exception of the N-terminus of Pks1 and the C-terminus of Pks9) contain sequences with convincing homology to previously-identified DDs[22,27] (the C-terminal DDs are referred to hereafter as ^{C}DDs and their partner N-terminal DDs as ^{N}DDs). By bioinformatics analysis, we were able to confidently assign the DDs acting at 6 of the 8 interfaces to the type '1a' class[22], and the remaining two sets of DDs as type '1b'[27] (**Supplementary Fig. 1**). In both cases, docking occurs between an α-helical ^{C}DD and a coiled-coil formed by the ^{N}DD, with specificity achieved via strategically-placed charge:charge interactions at the complex interface (**Supplementary Fig. 1**)[22,27].

Among the type 1a junctions, there were notably two which appeared compatible in terms of the translocated substrate: PKSs 3/4 + 7/8 and Pks 4/5 + 8/9 (**Supplementary Fig. 2**). Specifically, the functional groups at the critical α- and β- positions[14,28] of the transferred chains are identical at these junctions, and

correspondingly, the downstream KSs show similarities across several sequence motifs previously correlated with substrate specificity[14,20,29] (**Supplementary Fig. 2**). Targeting such interfaces thus allowed us, at least in principle, to overcome the functional block to the engineered systems represented by poor recognition of the incoming substrate by the directly downstream KS domain[21]. Ultimately, we targeted a new interface between Pks subunits 4 and 9 for two principle reasons. Firstly, Pks4 is at the origin of the structural variation between the stambomycin family members, and thus we anticipated that maintaining the subunit within the hybrid system would give rise to a corresponding series of truncated analogues, providing important evidence for their identity. Secondly, it was genetically more practical to modify the second set of interfaces due to splitting of the PKS subunits between two loci (**Fig. 1**).

To establish the novel junction, we initially modified the ^{C}DD of Pks4 (^{C}DD$_4$) to match that of Pks8 (the natural partner of the ^{N}DD of Pks9 (^{N}DD$_9$)), either by site-directed mutagenesis of residues previously identified as key mediators of interaction specificity (construct ^{C}DD$_4$ SDM; **Supplementary Fig. 3** and **Supplementary Table 1**)[22], or by exchange of the complete ^{C}DD docking α-helix of ^{C}DD$_4$ for that of ^{C}DD$_8$ (construct ^{C}DD$_4$ helix swap; **Supplementary Fig. 3** and **Supplementary Table 1**)[30]. Modifying the ^{C}DD$_4$ specificity 'code' to match that of ^{C}DD$_8$ required mutation of 3 residues, while for the ^{C}DD$_4$ helix swap, the terminal 16 amino acids of ^{C}DD$_4$ were exchanged for the corresponding 15 residues of ^{C}DD$_8$ (**Supplementary Fig. 3** and **Supplementary Table 1**). The genetic alterations were carried out in two distinct PKS contexts: (i) in the presence of the intervening subunits 5−8, which allowed for the possibility of competitive interactions between modified Pks4 and both Pks5 and Pks9; and ii) removing the intervening subunits 5−8, thus eliminating competition for binding of Pks4 by Pks5, and of Pks9 by Pks8 (**Supplementary Fig. 3**). We further generated a mutant in which Pks subunits 5−8 were deleted but no modification was made to ^{C}DD$_4$, in order to judge the intrinsic capacity of Pks4 and Pks9 to interact. Furthermore, genetic engineering was carried out in parallel by both PCR-targeting[31] and CRISPR-Cas9[32], in order to directly compare the efficacy of these two approaches, as well as evaluate the effect of the short scar sequence remaining in the chromosome following PCR-targeting.

Engineering the stambomycin PKS based on the classical module definition. The ^{C}DD$_4$ SDM and ^{C}DD$_4$ helix swap sequences were introduced in parallel into the *S. ambofaciens* genome using PCR targeting and CRISPR-Cas9 (full experimental details are provided in the **Supplementary Methods**). As discussed previously, the modifications were made both in the presence of the intervening subunits Pks5−8 and in their absence (**Supplementary Fig. 3**). As previous work has shown that production from the stambomycin biosynthetic gene cluster requires activation by constitutive overexpression of a pathway-specific LAL (Large ATP-binding regulators of the LuxR family) regulator[23], we additionally introduced the LAL overexpression plasmid (pOE484) into each of the mutants, using the empty parental plasmid (pIB139[33]) as a control. In total, this strategy resulted in 20 targeted strains harboring interface mutants (where K7N

refers to PCR targeting and CPN to CRISPR-Cas9 engineering): K7N1/pIB139, K7N1/OE484, K7N2/pIB139, K7N2/OE484, K7N3/pIB139, K7N3/OE484, K7N4/pIB139, K7N4/OE484, K7N5/pIB139, K7N5/OE484, K7N6/pIB139, K7N6/OE484, CPN1/pIB139, CPN1/OE484, CPN2/pIB139, CPN2/OE484, CPN4/pIB139, CPN4/OE484, CPN5/pIB139, CPN5/OE484 (**Table 1**, **Supplementary Tables 2**–**4**; despite extensive efforts the CPN3 mutant strain was not obtained). The principal difference between the K and CPN series of constructs is the presence of a 33 bp 'scar' sequence between the modified *pks4* and *pks9* genes (**Supplementary Fig. 4**). Construct K7N6 was assembled specifically to test the effect of this region, without any further modification to $^{C}DD_4$ and the intervening *pks5*–*pks8* genes.

Table 1 Summary of various strains generated, as well as the novel metabolites detected.

Strain	Modifications introduced	Stambomycins 1	Novel metabolites
ATCC/OE484	Wild type	✓	n.d.
K7N6/OE484[a]	33 bp scar[c]	✓	n.d.
K7N5/OE484	$^{C}DD_4$ helix swap, 33 bp scar	×	**2, 3, 4, 5**
CPN5/OE484[b]	$^{C}DD_4$ helix swap	✓	n.d.
K7N4/OE484	$^{C}DD_4$ site-directed mutagenesis (SDM), 33 bp scar	✓	n.d.
CPN4/OE484	$^{C}DD_4$ SDM	✓	n.d.
K7N3/OE484	Δ*pks5–8,* 33 bp scar	×	**2, 3, 4, 5**
K7N2/OE484	$^{C}DD_4$ helix swap, Δ*pks5–8,* 33 bp scar	×	**2, 3, 4, 5**
CPN2/OE484	$^{C}DD_4$ helix swap, Δ*pks5–8*	×	**2, 3, 4, 5**
K7N1/OE484	$^{C}DD_4$ SDM, Δ*pks5–8,* 33 bp scar	×	**2, 3, 4, 5**
CPN1/OE484	$^{C}DD_4$ SDM, Δ*pks5–8*	×	**2, 3, 4, 5**
ATCC/OE484/Pks4+TEI	TEI fused to Pks4	×	**2, 3, 4, 5**
CPN2/OE484/TEI SDM	TEI inactivation (Ser → Ala), $^{C}DD_4$ helix swap, Δ*pks5–8*	×	**2, 3, 4, 5**
CPN2/OE484/TEII SDM	TEII inactivation (Ser → Ala), $^{C}DD_4$ helix swap, Δ*pks5–8*	×	**2, 3, 4, 5**
CPN2/OE484/Δ478	Δ*samR0478,* $^{C}DD_4$ helix swap, Δ*pks5–8*	×	**2, 3, 4, 5**
CPN2/OE484/Δ479	Δ*samR0479,* $^{C}DD_4$ helix swap, Δ*pks5–8*	×	**6, 7, 8, 9**
CPN2/OE484/ACP_{13} SDM	ACP_{13} H1 modified[d], $^{C}DD_4$ helix swap, Δ*pks5–8*	×	**2, 3, 4, 5, 11**
ATCC/OE484/hy59_S1	$^{N}DD_9$+KS_{21} replaced by $^{N}DD_5$+KS_{14}, Δ*pks5–8*	×	**2, 3, 4, 5, 10, 11, 12**
ATCC/OE484/hy59_S2	$^{N}DD_9$+KS_{21} replaced by $^{N}DD_5$+$KS_{14/21}$, Δ*pks5–8*	×	**2, 3, 4, 5, 10, 11, 12, 14**
ATCC/OE484/hy59_S1/Δ479	Δ*samR0479*, $^{N}DD_9$+KS_{21} replaced by $^{N}DD_5$+KS_{14}, Δ*pks5–8*	×	**6, 7, 8, 9, 13**
ATCC/OE484/hy59_S2/Δ479	Δ*samR0479*, $^{N}DD_9$+KS_{21} replaced by $^{N}DD_5$+$KS_{14/21}$, Δ*pks5–8*	×	**6, 7, 8, 9, 13, 15**
ATCC/OE484/hy59_S1/ACP_{21} region swap	ACP_{21} L1+H2 modified[e], $^{N}DD_9$+KS_{21} replaced by $^{N}DD_5$+KS_{14}, Δ*pks5–8*	×	**2, 3, 4, 5**
ATCC/OE484/hy59_S1/ACP_{21} GtoD	ACP_{21} L1 modified[f], $^{N}DD_9$+KS_{21} replaced by $^{N}DD_5$+KS_{14}, Δ*pks5–8*	×	**2, 3, 4, 5, 11, 14**

[a]K7N (pronounced "cassette number") refers to mutants generated by PCR-targeting.
[b]CPN refers to mutants generated using CRISPR-Cas9.
[c]Use of PCR-targeting technique introduced a 'scar' sequence between *pks4* and *pks9*; for details, see Supplementary Fig. 4.
[d]H1 modified refers to mutation of six residues within the helix α1 region of ACP_{13} (EADQRR → PSERRQ); for details, see Supplementary Fig. 9.
[e]L1+H2 modified refers to exchange of the loop 1+helix α2 region of ACP_{21}; for details, see Supplementary Fig. 11.
[f]L1 modified indicates that one residue within the loop 1 region of ACP_{21} was mutated (G_{1499} → D); for details, see Supplementary Fig. 11.
n.d. indicates no novel metabolites were detected.

With the exception of K7N3, CPN4 and CPN5, extracts of the engineered mutant strains harboring pOE484 were analyzed by high performance liquid chromatography heated electrospray ionization high-resolution mass spectrometry (HPLC-HESI-HRMS) on a Dionex UltiMate 3000 HPLC coupled to a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer, and compared to extracts of the control strain containing pIB139[33] as well as the wild type *S. ambofaciens* containing pOE484, using SIEVE 2.0 screening software. K7N3, CPN4 and CPN5 were analyzed subsequently, and the data inspected manually.

Yield quantification was carried out with reference to a calibration curve generated with purified stambomycins A/B **1** (the limit of detection was found to be between 10 and 1 μg L^{-1}, and so any yields <10 μg L^{-1} must be considered an estimate). Novel metabolites not present in the control strains, and for which we obtained reliable exact masses, are listed in **Table 1** and **Supplementary Fig. 5**.

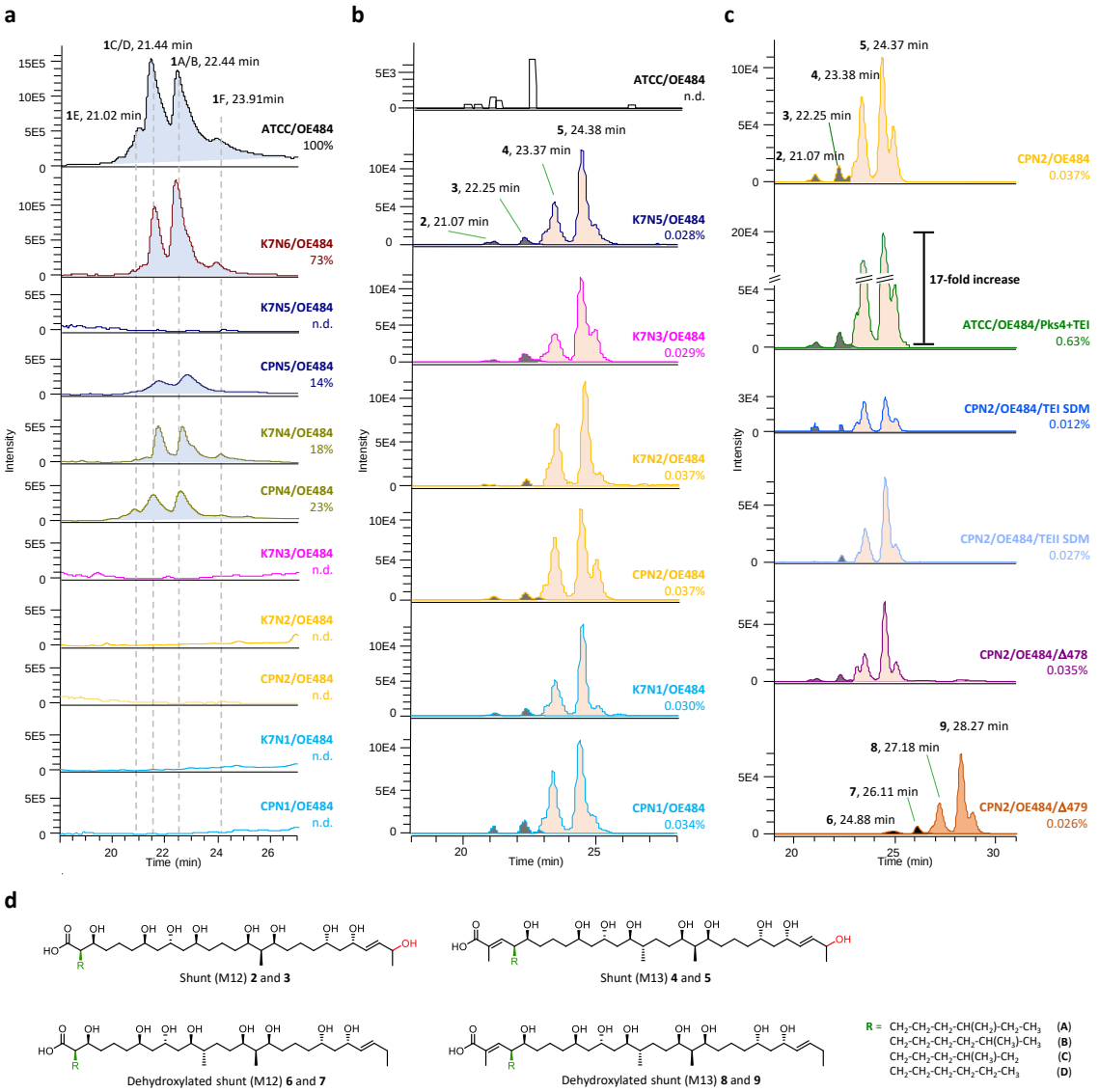


Fig. 3 Analysis of metabolites derived from PKS engineering based on the classical module definition. a HPLC-PDA analysis at λ_{max} 238 nm of stambomycins **1** from the wild type strain and various mutants. Quantification of all derivatives (**Supplementary Table 5**) was based on comparison to the yields of the wild type stambomycins **1** as determined using a calibration curve (**Supplementary Fig. 13**) (30 mg L^{-1} total yield of stambomycins A/B and C/D, set to 100%). **b** LC-ESI-HRMS analysis of mutants in which **1** was absent revealed a serious of shunt products (**2**–**5**) (the average yields (two measurements) relative to **1** in the wt are indicated). Shown are the extracted ion chromatograms (EICs) of **2**–**5**, using the calculated *m/z* values shown in **Supplementary Fig.5**. **c** LC-ESI-HRMS analysis of several CPN2-derived mutants (the yields of shunt products **2**–**5** are shown relative to **1** in the wt (average of four measurements)). Notably, the combined yield of **2**–**5** in mutant ATCC/OE484/Pks4+TEI was 17-fold higher than that from CPN2/OE484. A series of new compounds **6–9** was generated in strain CPN2/OE484 in which the gene *samR0479* was deleted. **d**

Chemical structures of shunts **2–9**. The structural differences among them are highlighted (green = R group; red = hydroxyl). Shunt products **2**, **4**, **6** and **8** correspond to stambomycin C/D derivatives, and **3**, **5**, **7** and **9** to stambomycin A/B derivatives. M12 and M13 indicate shunt compounds released from modules 12 and 13, respectively.

The first notable result is that the K7N6/OE484 mutant yielded a similar metabolic profile to *S. ambofaciens* wt (22 ± 3 mg L^{-1}, 73% relative yield (**Supplementary Table 5**)), showing that the scar sequence negatively impacted stambomycin production, but not dramatically (**Fig. 3**). By contrast, no stambomycins were observed, as anticipated, in all constructs in which Pks5−Pks8 had been removed (K7N1−3; CPN1,2) (**Fig. 3**). Stambomycins **1** were present, however, in strains K7N4 and CPN4 harboring $^{C}DD_4$ site-directed mutations and in the $^{C}DD_4$ helix swap strain CPN5, all of which still contained Pks5−Pks8, albeit at reduced amounts relative to the wild type (18%, 23% and 14% of wt, respectively) (**Fig. 3** and **Supplementary Table 5**). (Surprisingly, the metabolic profile of K7N5 reproducibly differed from that of CPN5, as no stambomycin-related metabolites were detected (**Fig. 3**)). These data suggested that while the mutations introduced into $^{C}DD_4$ reduced interaction with $^{N}DD_5$, they were not sufficient to disrupt natural chain transfer between Pks4 and Pks5, arguing that DD engineering to alter partner choice should be accompanied by removal of competing intersubunit interactions.

We did not find any evidence in the DD engineering experiments for any of the target 37-membered metabolites (**Supplementary Figs. 3** and **5**). However, all strains in which stambomycin production was abolished (**Table 1**) exhibited four new peaks in common (**Fig. 3b** and **Supplementary Fig. 5**) (other peaks corresponding to potentially novel compounds were observed, but none were shared between multiple strains). The determined exact masses and mass spectra (as exemplified by strain CPN2/OE484, **Fig. 3b**) correspond to truncated derivatives of stambomycins A/B and C/D respectively, following premature release from modules 13 and 12 of Pks4 (compounds **2**–**5**, **Fig. 3d** and **Supplementary Fig. 5**; ca. 8-fold greater yield of the module 13 products (**Supplementary Table 5**)). Further support for the identity of these shunt compounds was obtained by grafting the chain-terminating (type I) thioesterase (TE) domain from the C-terminal end of Pks9 to the C-terminus of Pks4 in order to force chain release at this stage. Indeed, identical compounds were produced, but at 17-fold increased yield relative to CPN2/OE484, consistent with active off-loading of the chains (**Fig. 3c**, **Supplementary Fig. 6** and **Supplementary Table 5**).

Based on their masses, both sets of shunt metabolites were hydroxylated on a single carbon, while none were found to bear the β-mycaminose of the mature stambomycins, consistent with the absence of the tetrahydropyran moiety to which it is normally tethered. To determine the location of the hydroxylation and therefore the hydroxylase responsible, we inactivated in mutant CPN2/OE484 the genes *samR0478* and *samR0479* encoding respectively, the stambomycin C-28 and C-50 cytochrome P450 hydroxylases[34]. While extracts of CPN2/OE484/Δ478 were unchanged relative to CPN2/OE484 (i.e. the hydroxyl group was still present), the CPN2/OE484/Δ479 mutant exhibited four new peaks with masses corresponding to the dehydroxylated shunt products (**Fig. 3**, **Supplementary Fig. 7** (compounds **6**–**9**) and **Supplementary Table 5**). Taken together, these data show that the unusual on-line modification catalyzed by SamR0479[34] which

is necessary for macrocyclization, occurs prior to chain extension by Pks5. While SamR0478 has also been speculated to act during chain assembly[34], hydroxylation evidently occurs downstream of Pks4, at least. The intriguing substrate structural and/or protein-protein recognition features controlling the timing of hydroxylation by these P450 enzymes remain to be elucidated.

Role of TE domains in release of the shunt metabolites. We attributed the observed shunt metabolites **2**–**5** to the lack of productive chain translocation between Pks4 and Pks9, causing intermediates to accumulate on ACPs 12 and 13. To evaluate whether these were released by spontaneous hydrolysis or enzymatically, we further investigated the role of Pks9 TEI[34] in chain release, as well as that of SamR0485, a proof-reading type II TE[35] located in the cluster. Both TEs were disabled by site-directed mutagenesis of the active site serines (Ser → Ala) (**Supplementary Fig. 6**).

Interestingly, inactivation of both the type I and type II TEs reduced the yields of shunt products **2**–**5** relative to the parental strain CPN2/OE484 (by 66% and 27%, respectively; average of duplicate experiments) (**Supplementary Fig. 6** and **Supplementary Table 5**). These data clearly show that premature release of the chains is catalyzed, at least in part, by both TEs in the cluster, although spontaneous liberation also occurs. While type II TEs typically interact with acyl-ACPs *in trans* to release blocked chains[35], the effect of the Pks9 TEI is less readily explained. One possibility is that the new productive docking interaction between Pks4 and Pks9 allows Pks9 to adopt an alternative conformation from which the TE can off-load intermediates bound to ACPs 12 and 13 of Pks4 (**Supplementary Fig. 6**). Although this mechanism is reminiscent of that used by the pikromycin PKS to generate both 12- and 14-membered rings[36], the pikromycin TEI is separated from its alternative ACP target by a single module, while Pks9 TEI is located five or four modules downstream from ACPs 12 and 13 in the engineered system, which would seem to necessitate substantial inter-subunit acrobatics.

Understanding the docking domain engineering via studies *in vitro* with recombinant domains. To better understand the results of the DD engineering, we studied *in vitro* the wild type DD pairs ($^{C}DD_4/^{N}DD_5$ and $^{C}DD_8/^{N}DD_9$), as well as binding between the modified versions of $^{C}DD_4$ and wild type $^{N}DD_9$. Design of suitable expression constructs in *E. coli* (**Supplementary Tables 1**–**3**) was based on bioinformatics analysis of the C-terminal ends of Pks4 and Pks8, and the N-termini of Pks5 and Pks9, and secondary structure analysis using PSIPRED[37] (**Supplementary Fig. 8**). Overall, we expressed and purified the following proteins in recombinant form from *E. coli*: $^{C}DD_4$ wt, $^{C}DD_4$ SDM, $^{C}DD_4$ helix swap, $^{N}DD_5$, and $^{C}DD_8$ (**Supplementary Fig. 8**, **Supplementary Table 4**). As $^{N}DD_9$ proved insoluble when expressed in *E. coli*, two versions with alternative start sites were obtained as synthetic peptides (Met and Val; **Supplementary Fig. 8**, **Supplementary Table 4**). Analysis of the individual ^{C}DDs by circular dichroism (CD) confirmed their expected high α-helical content ($^{C}DD_4$ wt (100 μM): 58%; $^{C}DD_8$ wt (100 μM): 49%), and showed no evident effect of the introduced mutations

on secondary structure (**Supplementary Fig. 8**). All of the constructs were further confirmed to be homodimeric by size exclusion chromatography-multi-angle light scattering (SEC-MALS) (**Supplementary Fig. 8**).

The two ^{N}DDs also exhibited α-helical character, though less pronounced than the ^{C}DDs ($^{N}DD_5$ (100 μM): 27%; $^{N}DD_9$ Met (100 μM): 21%; $^{N}DD_9$ Val (100 μM): 25%), and were monomeric by SEC-MALS (**Supplementary Fig. 8**). The latter result was surprising, as type 1a ^{N}DDs classically form a homodimeric coiled-coil domain (**Fig. 1**, **Supplementary Fig. 1**), but we recently identified functional, monomeric type 1 ^{N}DDs[38]. Indeed, we detected binding between the native pairs by isothermal titration calorimetry (ITC), with affinities in the range of those determined previously for matched pairs of DDs[27,38–40] ($^{C}DD_4$ + $^{N}DD_5$, K_d = 14.5 ± 0.9 μM; $^{C}DD_8$ + $^{N}DD_9$ Met, K_d = 33 ± 2 μM; $^{C}DD_8$ + $^{N}DD_9$ Val, K_d = 22 ± 1 μM) (**Supplementary Fig. 8**). Thus, while stable homodimerization of the ^{N}DDs may depend on the presence of a downstream homodimeric KS domain, their monomeric character did not preclude interaction with their ^{C}DD partners. Based on the higher affinity of the interaction, we could identify the $^{N}DD_9$ Val as the physiologically relevant construct. The observed binding stoichiometry (1 homodimeric ^{C}DD:2 monomeric ^{N}DDs), is consistent with the known structure of a type 1a complex in which two monomers of each DD are present (**Fig. 1**, **Supplementary Fig. 1**)[22]. As expected, no non-specific interaction was detected between native $^{C}DD_4$ and $^{N}DD_9$, explaining the lack of productive interaction between unmodified subunits Pks4 and Pks9 when the intervening multienzymes are deleted (strain K7N3) (**Fig. 3a**).

Analysis by ITC of binding between $^{C}DD_4$ SDM or $^{C}DD_4$ helix swap and $^{N}DD_5$ revealed the complete absence of interaction (**Supplementary Fig. 8**), and therefore that the introduced modifications were sufficient to disrupt communication between the native pair. Thus, the continued production of stambomycins **1** by K7N4, CPN4 and K7N5 harboring Pks5−Pks8 must be due to additional contacts between Pks4 and Pks5 beyond the docking domains, likely including the compatible ACP_{13}/KS_{14} interface. On the other hand, no interaction was detected between $^{C}DD_4$ SDM and $^{N}DD_9$, showing that this limited number of mutations was inadequate to induce productive contacts. This result is fully in accord with the absence of the expected mini-stambomycin products from these strains (K7N1/CPN1, **Fig. 3a**). By contrast, the $^{C}DD_4$ helix swap exhibited essentially the same binding to $^{N}DD_9$ Val as $^{C}DD_8$ (K_d = 21.0 ± 0.3 μM), demonstrating that exchange of just this helix is sufficient to redirect docking specificity[30]. Thus, inefficient docking is not at the origin of the failure of the $^{C}DD_4$ helix swaps to yield chain-extended products *in vivo* (strains K7N2/CPN2, **Fig. 3a**). We could therefore conclude at this stage that the problem arose from the non-native interface generated between ACP_{13} and KS_{21}, poor acceptance by KS_{21} of the incoming substrate during chain transfer and/or chain extension, and/or low activity towards the modified chain of domains/modules acting downstream.

Attempted optimization of the stambomycin DD mutants. We aimed next to improve the novel Pks4/Pks9 intersubunit interface in strain CPN2 ($^{C}DD_4$ helix swap + deletion of Pks5−8) by targeting helix αI of ACP_{13}, as the first 10 residues of this helix have been implicated previously in governing the interaction with the downstream KS domain at hybrid junctions[26]. Notably, multiple sequence alignment of all ACPs in the stambomycin PKS located at intersubunit junctions, revealed a unique sequence for each ACP in the helix αI region, consistent with a recognition 'code' for the KS partner, and the idea that mismatching these contacts might hamper productive chain transfer (**Supplementary Fig. 9**). Indeed, as mentioned previously, even when docking is interrupted, contacts between ACP_{13} and KS_{14} are apparently sufficient to enable chain translocation between Pks4 and Pks5 (**Fig. 3a**). In addition, an analogous strategy of optimizing the ACP_n/KS_{n+1} chain transfer interface was shown recently to substantially improve interaction between an ACP (JamC) derived from the jamaicamide B biosynthetic pathway, and the first chain extension module of the lipomycin PKS (LipPKS1)[41].

In our case, the first six residues of ACP_{13} helix αI were modified using CRISPR-Cas9 (EADQRR → PSERRQ), so that the full 10-residue recognition sequence matched that of ACP_{20}, the natural partner of KS_{21} (**Supplementary Fig. 9**). Analysis of extracts of the resulting strain CPN2/OE484/ACP13 SDM by HPLC-MS revealed at best minute amounts (highest yield of 0.5 μg mL^{-1}) of target cyclic mini-stambomycins A/B (**11**), lacking the hydroxyl group introduced by SamR0478 (**Fig. 4** and **Supplementary Fig. 9**). Thus, while this experiment finally yielded the first evidence for successful chain transfer between Pks4 and Pks9 followed by subsequent chain extension by Pks9 and TE-catalyzed release, the overall efficiency of the system remained poor. Interestingly, however, the yields of the four shunt metabolites **2**–**5** were as much as 48-fold higher from the ACP_{13} helix swap mutant than from CPN2/OE484, showing that improved interactions between ACP_{13} and KS_{20} facilitated release of the stalled intermediates from ACPs 12 and 13, presumably via remote action by the TEI domain.

Engineering mini-stambomycins by maintaining the native ACP_{13}/KS_{14} junction (alternative module definition). Cumulatively, the results obtained with the docking domain engineering identified KS_{21} as one potential bottleneck in the engineered PKS. Our parallel strategy based on the alternative module definition (**Fig. 2**) allowed us to directly test this idea. Specifically, we investigated the effects of preserving the native $^{C}DD_4/^{N}DD_5$ pair and either the majority of KS_{14}, or a little more than half of the domain, resulting in a KS_{14}/KS_{21} hybrid. For this, we used two different splice sites in KS_{14}: i) at the end of the domain in a highly-conserved region (GTNAHV) exploited recently to efficiently swap downstream AT domains[42]; and, ii) at a site corresponding to a recombination hot spot identified during induced evolution of the rapamycin (RAPS) PKS[43], yielding the KS_{14}/KS_{21} chimera (**Fig. 4** and **Supplementary Fig. 10**). Both of these modifications were introduced into *S. ambofaciens* using CRISPR-Cas9, while simultaneously removing Pks5−Pks8, yielding

respectively after co-transformation with pOE484 and the control plasmid pIB139, strains ATCC/OE484/hy59_S1, ATCC/pIB139/hy59_S1, ATCC/OE484/hy59_S2, and ATCC/pIB139/hy59_S2.

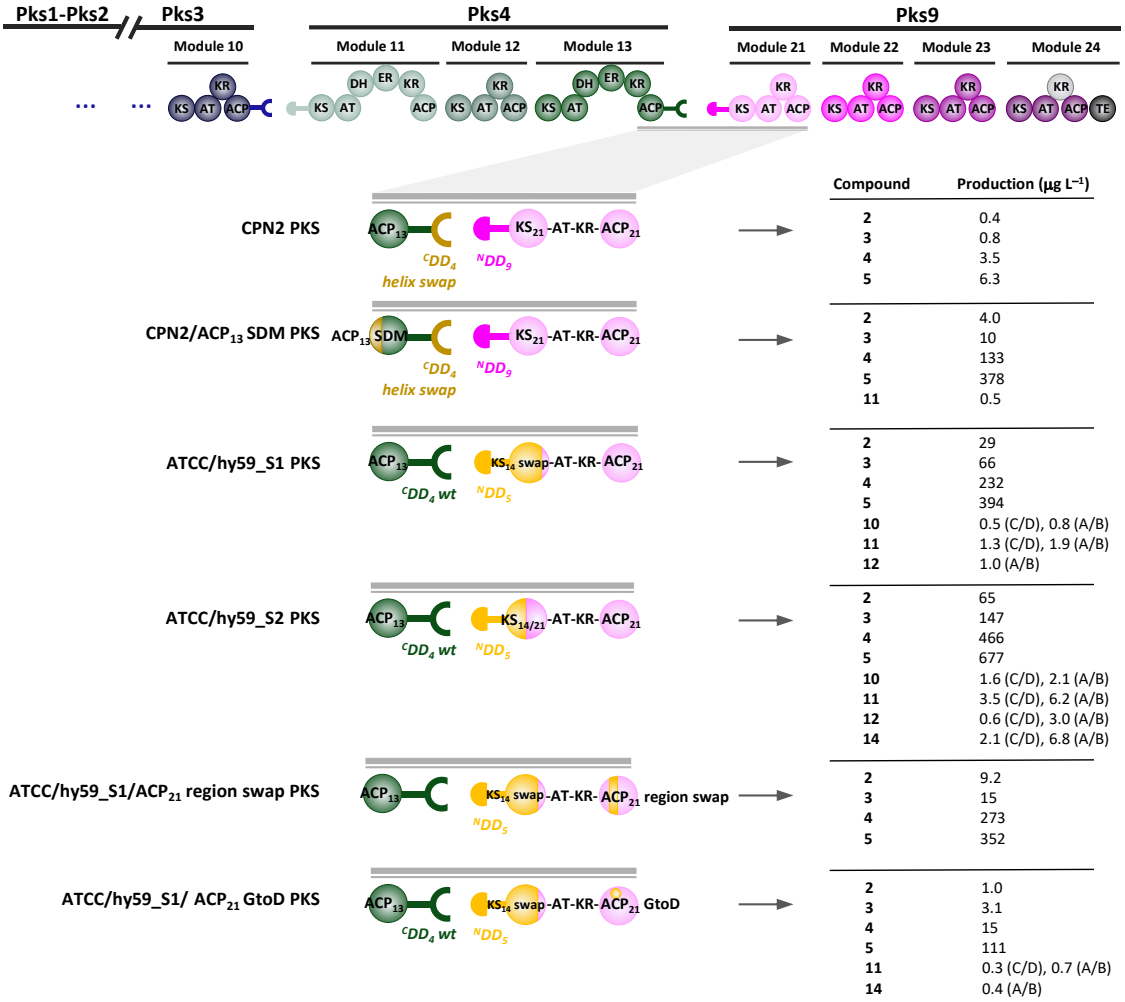


Fig. 4 Engineering of functional mini-stambomycin PKSs. The various strategies used in each case are represented schematically, along with the obtained products and their maximum yields (estimated for <10 μg L^{-1}) (full analysis of all constructs is provided in **Supplementary Table 5**). The engineering starting point, CPN2 PKS, contains a functional $^{C}DD_4$ helix swap/$^{N}DD_9$ docking interaction (swapped docking α-helix shown in dark yellow), but a mis-matched ACP_{13}/KS_{21} interdomain interaction. This PKS yielded only shunt products **2**–**5**. The CPN2/ACP_{13} SDM PKS, in which the ACP_{13} helix αI has been modified to match that of ACP_{20} (dark yellow ball), generates mini-stambomycin derivatives (**11**, a cyclic form lacking the internal hydroxy, **Supplementary Fig. 9**). The ATCC/hy59_S1 and S2 constructs were based on the alternative module definition, as the engineering point was selected downstream of the $^{C}DD_4/^{N}DD_5$ interface within the KS_{21} domain. Of the two junctions, that in which the fusion was located essentially at the midpoint of the domain (ATCC/hy59_S2) functioned better than that which included the majority of KS_{14} (ATCC/hy59_S1), although both PKSs led to successful generation of three novel mini-stambomycins, both as their free acids (**10**) and in macrolide form (**11** and **12**) (**Supplementary Fig. 10**). In an attempt to boost yields from ATCC/hy59_S1, two further constructs were created by modification of ACP_{21} – either by swapping a region implicated in KS/ACP communication during extension (ATCC/hy59_S1/ACP_{21} region swap), or by mutating a single key residue within the motif (ATCC/hy59_S1/ACP_{21} GtoD). The ATCC/hy59_S1/ACP_{21} region swap yielded only the shunt metabolites **2**–**5**, while solely **11** was detected from ATCC/hy59_S1/ACP_{21} GtoD, showing that the introduced changes did not work as intended. We observed in addition

from ATCC/hy59_S2 PKS and ATCC/hy59_S1/ACP_{21} GtoD, shunt product **14** corresponding to the chain released from module 21 (**Supplementary Figs. 9** and **10**), identifying the downstream module 22 as a blockage point.

Analysis of culture extracts revealed the presence in both ATCC/OE484/hy59_S1 and ATCC/OE484/hy59_S2 relative to the controls, of a novel series of 37-membered metabolites (**Fig. 4**). The measured masses were consistent with the desired mini-stambomycins either as their free acids or in cyclic form (metabolites **10**–**12**, **Fig. 4**). Signals corresponding to the A/B and C/D derivatives of all metabolites were detected, providing important evidence for their identities, as well as both the C-14 hydroxylated **12** and non-hydroxylated **11** forms of the cyclic mini-stambomycins (C-14 corresponds to C-28 in the parental compounds (**Fig. 1**)). It is not surprising that the corresponding E and F forms were not detected, as their yields even from the wild type are much lower than the A–D derivatives (**Fig. 3a**). The observation of non-hydroxylated **11** shows notably that internal hydroxylation by SamR0478 is not an absolute prerequisite for TE-catalysed macrolactonization, and argues that hydroxylation of the mini-stambomycins only takes place on the macrocyclic compound. Although compounds **11** and **12** incorporate the tetrahydropyran moiety of the parental stambomycins **1** which undergoes glycosylation, derivatives bearing β-mycaminose were not observed, presumably due to poor recognition of the overall modified macrocycle by glycosyl transferase SamR0481[23].

The yields of the target compounds were minor relative to wild type stambomycins (metabolites **10**, **11** and **12** from ATCC/OE484/hy59_S1 were obtained at highest yields of 0.3, 3.2 and 1.0 $\mu g\ L^{-1}$ (4.5 $\mu g\ L^{-1}$ total), respectively), but nonetheless approximately 4-fold higher from ATCC/OE484/hy59_S2 incorporating the hybrid KS_{14}/KS_{21} than from the full KS_{14} swap (3.7, 9.7 and 3.6 $\mu g\ L^{-1}$ (17 $\mu g\ L^{-1}$ total); 1500-fold lower yields than **1**) (**Fig. 4** and **Supplementary Fig. 10**). Although the low titers of these compounds precluded their structure elucidation by NMR, we obtained additional confirmatory evidence for their identities by inactivation of *samR0479* (which introduces the hydroxyl used for macrocyclization), which resulted in exclusive production of linear dehydroxy mini-stambomycins **13** (**Supplementary Fig. 10**).

As observed previously, the strains also produced substantial quantities of the shunt products **2**–**5** (inactivation of *samR0479* led correspondingly to the dehydroxy versions of these compounds **6–9** (**Supplementary Figs. 7** and **10**)). The yields were ca. 80-fold higher than those of the corresponding mini-stambomycins, with the highest titers observed in the strain incorporating the hybrid KS_{14}/KS_{21}. The amount of shunt metabolites was also approximately 123-fold higher than from strain CPN2/OE484 (which incorporates an ACP_{13}-$^{C}DD_4$ swap/$^{N}DD_9$-KS_{21} interface) (**Figs. 3a** and **4**, **Supplementary Table 5**). Thus, contrary to expectation, although using the KS as a fusion site improved communication between Pks4 and Pks9, it also substantially boosted TEI-mediated off-loading of stalled upstream intermediates.

In principle, such stalling could result from a slow rate of chain extension in the now hybrid acceptor module (for example, in the full KS swap construct, KS_{14} and ACP_{21} are completely mismatched for chain extension). To evaluate this idea, we modified ACP_{21} within ATCC/OE484/hy59_S1 incorporating the full-

length KS_{14}, targeting a sequence region previously identified as mediating intramodular communication between the KS and ACP during chain extension (**Supplementary Fig. 11**)[19,26]. Specifically, we exchanged loop 1 and the initial portion of helix αII of ACP_{21} for the corresponding sequence of ACP_{14}, using CRISPR-Cas9 (**Supplementary Fig. 11**). As we anticipated that creation of this substantially hybrid ACP might engender structural perturbation, we also engineered a minimal mutant of ACP_{21} in which only one of the two most critical residues in the recognition motif was mutated to the corresponding amino acid in ACP_{14} (G_{1499} of Pks9 → D; the second residue, R, of the motif is already common to the two ACPs) (**Supplementary Fig. 10**). Analysis of the loop/helix αII swap by HPLC-MS showed that all mini-stambomycin production had been abolished (**Supplementary Fig. 10**), consistent with the anticipated disruption to ACP_{14} structure. Production by the ACP site-directed mutant was not any better than by the full KS swap construct (**Fig. 4** and **Supplementary Fig. 10**), as only metabolite **11** remained above the limits of detection.

In principle, the hybrid KS_{14}/KS_{21} domain may have worked better than KS_{14} for chain extension due to improved interaction with ACP_{21}, with stalling displaced to later modules. If this were the case, we might expect to see accumulation in the medium of shunt metabolites corresponding to the intermediate generated by module 21. Indeed, in the case of strain hy59_S2 (chimeric KS_{14}/KS_{21}) but not hy59_S1 (KS_{14}), we detected masses consistent with the A/B and C/D forms of intermediate **14** generated by module 21, at yields comparable to those of the final mini-stambomycins (**Fig. 4**, **Supplementary Fig. 10**) (and correspondingly, **15**, the dehydroxylated analogue of **14**, was detected in the SamR0479 mutant (**Supplementary Fig. 10**)). The same metabolite **14** was identified from the ACP_{21} G →D mutant (**Fig. 4** and **Supplementary Fig. 11**), consistent with interrupted chain transfer to KS_{22}. Taken together, these data confirm module 22 as a new blockage point in the engineered systems.

Relative efficacy of PKS engineering using PCR-targeting and CRISPR-Cas9. As multiple of our core constructs were generated by both PCR-targeting and CRISPR-Cas9, we were able to directly compare the efficiency of the two techniques (**Fig. 3** and **Supplementary Fig. 4**). Globally, our results confirm that both approaches can be employed to introduce large-scale modifications to PKS biosynthetic genes (i.e. deletions of single or multi-gene regions)[32,44–46]. We have also demonstrated, for only the second time to our knowledge, that CRISPR-Cas9 can be leveraged to specifically modify modular PKS domains[47]. Of the two methods, CRISPR-Cas9 was the more rapid, as the corresponding constructs were engineered in approximately half of the time. In addition, while CRISPR-Cas9 allowed for direct modification of the host genome, PCR-targeting relied on the availability of suitable cosmids housing the target genes, and resulted in a 33 bp *attB*-like 'scar' sequence in the genome (**Supplementary Fig. 4**)[48]. In addition to hampering iterative use of the approach, the scar apparently provoked a moderate reduction in stambomycin yields in mutant K7N6 compared to the wild type, an effect also noted upon comparison of several analogous mutant strains (e.g. K7N4 vs. CPN4, **Fig. 3**). Nonetheless, we did encounter certain difficulties with use of CRISPR-

Cas9 (i.e. failure to obtain construct CPN3, occasional reversions to wild type, etc.), observations motivating ongoing efforts in other laboratories to further enhance the suitability of CRISPR-Cas9 for editing PKS pathways[47,49–54].

Discussion

In this work, we have utilized an approach based on the state-of-the-art in PKS engineering to modify the stambomycin PKS (**Fig. 5**). Specifically, we aimed to remove the four PKS subunits between Pks4 and Pks9 in the assembly line which together house seven chain extension modules, to generate a series of 37-membered 'mini-stambomycins'. While in principle such a change might have been possible by directly fusing Pks4 and Pks9 via a suitable intermodular linker, this approach would have resulted in a heptamodular subunit whose size is far in excess of the tetramodular multienzymes present in the system. We have also demonstrated recently the low efficacy of this strategy when the module downstream of the linker is N-terminal in its native subunit context (as with module 21 of Pks9)[18].

As an initial approach (**Fig. 5**), we modified $^{C}DD_4$ to render it compatible with $^{N}DD_9$, with the aim of inducing productive communication between Pks4 and Pks9, while leaving all modular units intact. This modified PKS relied for function on both a non-native chain transfer interface (ACP_{13}/KS_{21}), and the intrinsic tolerance of the downstream KS/modules to the incoming substrate. We were optimistic this experiment might work given the structural similarities between the native substrates of KS_{14} and KS_{21} at least directly adjacent to the acyl terminus, as well as the fact that the stambomycin PKS generates a small family of metabolites, and therefore must exhibit some intrinsic tolerance to structural variation. Although we showed *in vitro* with recombinant DD pairs that a docking helix-swapped mutant of $^{C}DD_4$ communicated effectively with $^{N}DD_9$, chain transfer across the engineered interface did not occur *in vivo*, as evidenced by the accumulation of multiple shunt products. While our attempt to render the ACP_{13}/KS_{21} junction more native by site-directed mutagenesis did result in certain target metabolites, the most significant effect was to increase the yields of the truncated chains.

Having narrowed down the biosynthetic block to events occurring downstream of the engineered junction, we next carried out interface engineering based on proposed alternative module boundaries, leveraging fusion points within the KS domain (**Fig. 5**). In this case, sites were selected to either maintain essentially the whole of KS_{14}, or to create a hybrid KS_{14}/KS_{21} domain. This strategy at once preserved key elements of the ACP_{13}/KS_{14} chain transfer junction, and in the case of the almost full-length KS_{14}, ensured that the domain had the appropriate substrate specificity for the incoming chain. Interestingly, the construct incorporating the chimeric KS functioned best, producing the desired mini-stambomycins in both linear and macrocyclic forms. These data identify this location in the middle of the KS[43] as a potentially general fusion site, perhaps because it preserves elements of both the chain transfer and chain extension interfaces with the two partner ACP domains (ACP_{13} and ACP_{21}, in this case). Intriguingly, a site within the

analogous condensation (C) domains of modular nonribosomal peptide synthetase (NRPS) systems, has also recently emerged as a useful fusion point for generating engineered hybrids[55].

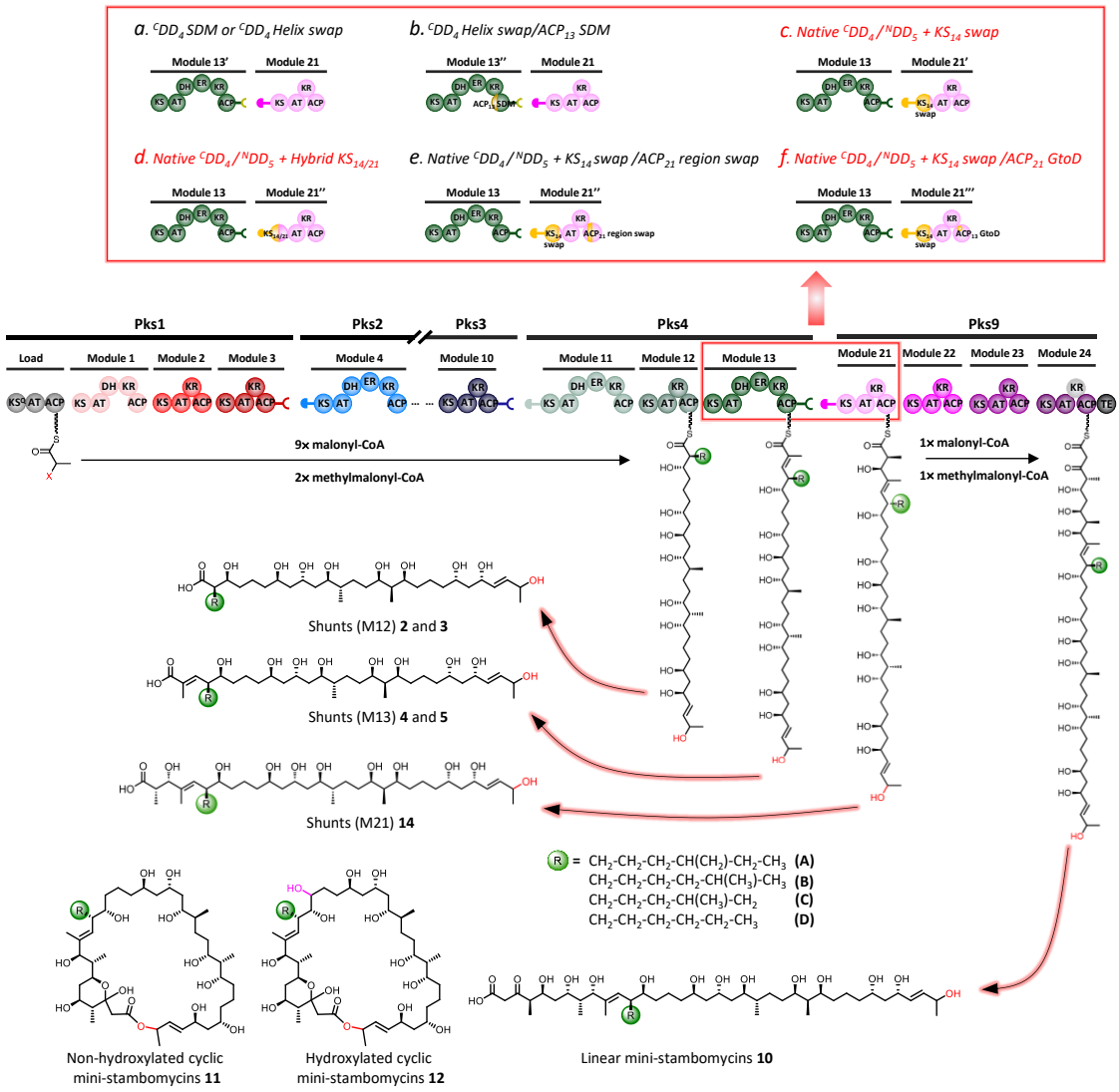


Fig. 5 Summary of the engineering strategies applied in this work to the stambomycin PKS. Inset are the six distinct approaches used, and the structures of the resulting metabolites are shown. The strategies giving rise to the target mini-stambomycins **10**–**12** are indicated in red. The hydroxyl group shown in pink is introduced by the P450-hydroxylase SamR0478, and that in red, by SamR0479.

Our results also showcase the intrinsically high tolerance of the Pks9 TEI domain towards shorter substrates. Indeed, the data also demonstrate that this TEI domain participates in off-loading the shunt metabolites from the upstream subunit, and that this activity interferes with passage of the chain to subsequent modules. Unfortunately, our attempts to boost yields of the mini-stambomycins by engineering the condensation interface between KS_{14} and ACP_{21} were unsuccessful, both when the full ACP_{21} recognition loop/helix αII region was swapped for that of ACP_{14}, and when a single site-directed mutation was made at

a putatively critical position (**Fig. 5**). This result is surprising in light of the beneficial effects reported *in vitro* of both of these modifications on chain extension carried out by mis-matched KS and ACP domains sourced from the erythromycin PKS (DEBS)[26]. Apparently, the introduced changes were not sufficient to ensure effective communication between the KS_{14} and ACP_{21} domains (or were in fact deleterious to function), and/or any benefit was masked by the poor tolerance of the downstream modules to the modified intermediates.

To fully judge the efficacy of this work, it is instructive to compare it to the other two examples in the literature in which full biosynthetic systems have been re-engineered to remove multiple internal modules[10]. In the first, recently-reported case, the neoaureothin (Nor) hexamodular PKS was 'morphed' into the evolutionarily-related aureothin (Aur) tetramodular PKS by removing the second bimodular subunit, NorA'. As in our work, the authors initially attempted to engineer a new interaction between the monomodular subunits NorA and NorB flanking NorA' using compatible docking domains, by exchanging the type 1b ^{N}DD of NorB for the type 1a ^{N}DD of NorA' (the natural partner of NorA). When the target metabolite was not obtained, they relocated the fusion site to the KS-AT linker downstream of the conserved KS region in NorB, thereby maintaining the native NorA ACP-^{C}DD/^{N}DD-KS NorA' junction. Ultimately, several linker variants had to be evaluated before a functional sequence was identified, in part by serendipity (indeed it is 1 residue longer than the native linker). Overall, the yields of the targeted chain-shortened metabolites were reduced approximately 18-fold compared to the parental neoaureothin (to ca. 2.5 mg L^{-1}), a much less significant penalty than engendered by our engineering strategy. Presumably, the superior titers obtained in this experiment reflect the much higher intrinsic amenability of the Nor PKS to conversion into an Aur PKS, as the Nor PKS likely evolved from an Aur PKS by subunit insertion[10]. Nevertheless, the newly-created NorA/NorB interface was also only partially functional, as product corresponding to the intermediate generated by iterative action of the upstream subunit NorA was still obtained.

The second relevant investigation concerns the accelerated evolution (AE) of the RAPS PKS, based on spontaneous induced homologous recombination between its component modules[43]. As mentioned earlier, several of the resulting systems incorporated intermodular fusion sites essentially at the mid-point of the respective KS domains, and so can be compared to our best performing construct hy59_S2. Notably, yields from the hybrid RAPS PKSs from which either 3 or 6 modules were removed, were reduced by a maximum of 33% relative to that of the parental compound. We propose two explanations for the higher functionality of these systems relative to hy59_S2. First of all, in every case, the module downstream of the newly-formed junctions in the contracted RAPS PKSs was internal to its respective subunit (unlike module 21 of Pks9), consistent with the idea that such modules boast intrinsically broader substrate specificity[18]. Secondly, the KS domains of the RAPS PKS exhibit unusually high mutual sequence identity (85−90%)[56]. This means, in effect, that the same KS domain acts on a large variety of substrates of differing length and functionality, contributing to the tolerance of all modules downstream of the fusion site. In contrast, the

KSs in the stambomycin PKS have substantially lower sequence identity (62−80%), and thus KSs 22−24 in hy59_S2 likely represent a specificity barrier to efficient transfer and extension of the modified intermediates.

Taken together, this set of results shows that contracting PKS systems represents a viable approach to accessing truncated polyketide derivatives of variable length, including macrocycles. Whether such systems are generated rationally or using an AE process, the most efficient hybrids will likely result: i) from PKSs whose modules (and in particular KS domains) exhibit a substantial degree of mutual sequence identity and thus intrinsically high substrate tolerance (or which can be adapted by mutagenesis to broaden their specificity[29]); and, ii) when novel junctions are created with downstream modules which are situated at internal positions within their subunits. The data also reinforce the idea that in cases where communication at modified interfaces occurs via non-covalent protein-protein interactions, at least a portion of the KS downstream from the docking domains should be included to boost efficiency[10,11,16]. Finally, our work has identified an increase in TEI-mediated proof-reading provoked by such interface engineering. Elucidating the mechanism underlying this unexpected intersubunit release activity, and thus how to effectively suppress it, should be a profitable avenue for further boosting product titers.

Methods

Bioinformatics analysis. To underpin the interface engineering strategy, the extremities of all the stambomycin PKS subunits were analyzed to identify the boundaries of the most C-terminal and N-terminal function domains (ACP and KS, respectively), and thus the regions potentially containing docking domains (DDs). The resulting sequences were compared by multiple sequence alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo)[57], to bonafide and putative DD sequences from multiple DD classes, including those present at the DEBS 2/DEBS 3 interface (type 1a, PDB ID:1PZQ, 1PZR[22]) and the PikAIII/PikAIV junction (type 1b, PDB ID: 3F5H[27]), to allow for type classification. To identify suitable boundaries for DD heterologous expression in *E. coli*, the secondary structure of the putative DD regions was predicted using PSIPRED 4.0 (http://bioinf.cs.ucl.ac.uk/psipred/)[37]. Analysis for potential specificity-conferring residues in the stambomycin PKS ketosynthase (KS) domains was carried out by multiple sequence alignment against model KS domains[14,20,29], using Clustal Omega[57].

General methods. All reagents and chemicals were obtained from Sigma-Aldrich, except the following: BD (tryptone, yeast extract, TSB powder), Thermo Fisher Scientific (Tris), VWR (glycerol, NaCl, $NaNO_3$), ADM, France (NutriSoy flour), and New England Biolabs (T4 DNA ligase, restriction enzymes). Oligonucleotide primers and two additional synthetic DNA fragments for CPN4 and CPN5 constructs were synthesized by Sigma-Aldrich (**Supplementary Table 2**). The docking domains $^{N}DD_9$ Val and $^{N}DD_9$ Met **Supplementary Table**

1) were obtained as synthetic peptides from GeneCust. DNA sequencing of PCR products was performed by Sigma-Aldrich and Eurofins.

PCR reactions were performed with Taq DNA polymerase (Thermo Fisher Scientific) or Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) when higher fidelity was required. Isolation of DNA fragments from agarose gel, purification of PCR products and extraction of plasmids were carried out using the NucleoSpin® Gel and PCR Clean-up or NucleoSpin® Plasmid DNA kits (Macherey Nagel, Hoerdt, France).

Strains and media. *E. coli* BL21 strains were obtained from Novagen. Unless otherwise specified, all *E. coli* strains were cultured in LB medium (yeast extract 10 g, tryptone 5 g, NaCl 10 g, distilled water up to 1 L, pH 7.0)[58] or on LB agar plates (LB medium supplemented with 20 g L^{-1} agar) at 37 °C. *Streptomyces ambofaciens* ATCC23877 and the derived mutants were grown in TSB (TSB powder 30 g (tryptone 17 g, soy 3 g, NaCl 5 g, K_2HPO_4 2.5 g, glucose 2.5 g), distilled water up to 1 L, pH 7.3) or on TSA plates (TSB medium supplemented with 20 g L^{-1} agar), and sporulated on SFM[59] agar plates (NutriSoy flour 20 g, D-mannitol 20 g, agar 20 g, tap water up to 1 L) at 30 °C. All strains were maintained in 20% (*v/v)* glycerol in 2 mL Eppendorf tubes and stored at −80 °C.

For fermentation of *S. ambofaciens* ATCC23877 and its mutants, spores were streaked on TSA with appropriate antibiotics and after incubation 48 h at 30 °C, a loop of mycelium was used to inoculate 7 ml of MP5 medium (yeast extract 7 g, NaCl 5 g, $NaNO_3$ 1 g, glycerol 36 mL, MOPS 20.9 g, distilled water up to 1 L, pH 7.4) supplemented with selective antibiotics and sterile glass beads, followed by incubation at 30 °C and 200 rpm for 24−48 h. Finally, the seed culture was centrifuged and resuspended into 2 mL fresh MP5 before being inoculated into 50 mL MP5 medium in a 250 mL Erlenmeyer flask, and cultivated at 200 rpm and 30 °C for 4 days.

PCR-targeting-based genetic engineering (exemplified by mutant ATCC/OE484/K7N1, **Supplementary Fig. 4**). To render the BAC BAA9ZA8 proficient for selection following conjugation, its chloramphenicol resistance gene was replaced using a PCR-targeting approach, by a kanamycin resistance gene cassette sourced from pIJ776[31], resulting in BAC1 (**Supplementary Table 3**). The cassette *attL+aac(3)IV+oriT+attR* was amplified from the plasmid pSPM88T[60] (**Supplementary Fig. 4**) using primers 9996 and 9997 (**Supplementary Table 2**), affording PCR amplicon *PCR-K7N1.* The PCR fragment was then electro-transformed into *E. coli* BW25113/pKD20/BAC1[61], giving rise to mutant BAC1_K7N1 (in which the C-terminus of *pks4* was modified and the genes *pks5−pks8,* were deleted)[31]. The BAC1_K7N1 was then introduced into *E. coli* ET12567/pUZ8002[62] and then transferred to *S. ambofaciens* wild type *via* intergeneric conjugation. The resulting exconjugants (ATCC/K7N1_*aac(3)IV+oriT*) were selected for their apramycin resistance and kanamycin sensitivity (i.e. a phenotype consistent with successful double cross-over). The correct mutations were confirmed by PCR and sequencing. Subsequently, the disruption cassette was

excised using the excisionase and integrase of pSAM2 encoded by pOSK111 as described[60], leaving a 33 bp 'scar' sequence (mutant ATCC/K7N1). Successful removal of the cassette was verified by PCR and DNA sequencing Finally, the LAL regulator overexpression plasmid pOE484[23] or the parental vector pIB139[33] was introduced into the strain giving rise to mutants K7N1/OE484 and K7N1/pIB139, respectively. The same overall procedure was applied to construct mutants K7N2/OE483, K7N2/pIB139, K7N3/OE483, K7N3/pIB139, K7N4/OE483, K7N4/pIB139, K7N5/OE483, K7N5/pIB139, K7N6/OE483 and K7N6/pIB139. An analogous PCR-targeting approach (**Supplementary Fig. 4**) was also employed to inactivate *samR0478* and *samR0479* using appropriate BACS[34] (**Supplementary Table 3**).

CRISPR-Cas9-mediated genetic engineering. Plasmids pCRISPomyces-2 (and associated cloning and screen protocols)[32] (used for construction of all mutants except ATCC/hy59_S1 and ATCC/hy59_S2) and pCRISPR-Cas9[45], respectively, were used for CRISPR-Cas9-based genome editing, as described previously. The two systems differ in the way in which Cas9 is expressed; in the case of pCRISPomyces-2, the nuclease is expressed constitutively, while in the pCRISPR-Cas9 system, its expression is under inductive control by thiostrepton (Tsr). The crRNA sequence was selected to match the DNA segment which contains NGG on its 3' end (N is any nucleotide, and the NGG corresponds to the protospacer-adjacent motif (PAM). The annealed crRNA fragment and two homologous arms (HAL and HAR, flanking the target region) were sequentially inserted into the delivery plasmid pCRISPomyces-2 using the respective restriction sites *Bbs*I and *Xba*I, to afford the specific recombinant plasmid for each mutant (**Supplementary Fig. 4**). Correspondingly, an sgRNA cassette (tracrRNA+sgRNA) and two homologous arms were inserted into the plasmid pCRIPR-Cas9 using sites *Nco*I, *SnaB*I and *Stu*I, respectively (**Supplementary Fig. 10**). In addition, the crRNA was designed to be located within the region to be deleted (**Supplementary Fig. 4**) to avoid Cas9-catalyzed cleavage occurring in the genome of the resulting mutant. In the case of site-directed mutants, additional DNA fragments containing the targeted mutations were inserted between the two homologous arms. In addition, the DNA sequence with the fragments identical to the crRNA was modified, so as to avoid subsequent Cas9-catalyzed cleavage of the obtained mutants (**Supplementary Fig. 9** and **11**).

Overexpression and purification of docking domains. The wild-type docking domains ($^{C}DD_4$, $^{N}DD_5$, $^{C}DD_8$, $^{N}DD_9$ Val and $^{N}DD_9$ Met) and mutant docking domains ($^{C}DD_4$ SDM, $^{C}DD_4$ helix swap) were amplified from genomic DNA of *S. ambofaciens* wild type and the relevant mutants, using forward and reverse primers incorporating *Bam*HI and *Hin*dIII restriction sites, respectively (**Supplementary Table 2**). The PCR amplicons were digested using FD *Bam*HI and FD *Hin*dIII, and then ligated into the equivalent sites of vector pBG-102 (Center for Structural Biology, Vanderbilt University). In the case of all ^{C}DDs which lacked aromatic residues, a tyrosine residue (codon TAT incorporated in the forward primer, **Supplementary Table 1**) was added at the N-terminal ends (so as not to interfere with docking with the ^{N}DD partner) to allow efficient monitoring

by UV-Vis during the purification, as well as reliable measurement of protein concentration necessary for binding studies by ITC.

The resulting constructs pBG102-$^{N}DD_5$, pBG102-$^{C}DD_8$ and pBG102-$^{N}DD_9$ were used to transform *E. coli* BL21 (DE3). For $^{C}DD_4$ and its mutants, these were transformed into Rosetta™ 2(DE3), as these constructs contain 8 codons rarely used in *E. coli*. Positive transformants were selected on LB agar supplemented with kanamycin (50 μg mL^{-1}) (25 μg mL^{-1} chloramphenicol was also added for expression in Rosetta™ 2(DE3)). A single colony was transferred to LB (10 mL) supplemented with antibiotics, and the culture grown at 37 °C and 200 rpm for overnight. The 1 mL overnight culture was used to inoculate LB media (1 L) supplemented with appropriate antibiotics, and then incubated at 37 °C and 200 rpm to an optical density of 0.8, at which point protein synthesis was induced by the addition of IPTG (final concentration 0.1 mM). After incubation at 18 °C and 200 rpm for 18 h, cells were collected by centrifugation at 8000*g* for 30 min, resuspended in 40 mL protein purification buffer A (50 mM Tris-HCl, 400 mM NaCl, 10 mM imidazole, pH 8.0), and lysed by sonication. Following centrifugation at 20000*g* and filtration using a 0.45 μm membrane, the soluble cell lysates were loaded onto 2 × 5 mL HisTrap HP (GE Healthcare) columns (two 5 mL columns in series) equilibrated in buffer A, and purified by preparative protein purification chromatography using an ÄKTA Avant system. The following program was applied: sample loading, 1 mL min^{-1}; washing, 2 mL min^{-1}, 10 column volumes of buffer A; elution, 2 mL min^{-1}, 5 column volumes of buffer B (50 mM Tris-HCl, 400 mM NaCl, 250 mM imidazole, pH 8.0); elution, 2 mL min^{-1}, 2 column volumes of buffer C (50 mM Tris-HCl, 400 mM NaCl, 500 mM imidazole, pH 8.0).

All His_6-SUMO-tagged proteins were collected (fractions containing the protein of interest were selected based on the UV chromatography and SDS-PAGE gel), and transferred into dialysis bag containing His_6-tagged human rhinovirus 3C protease (H3C) (1−2 μM). The dialysis bag was then placed into a container filled with buffer D (50 mM Tris-HCl, 400 mM NaCl, pH 8.0), and the cleavage allowed to proceed at 4 °C overnight. The resulting proteins, which incorporated a non-native N-terminal GPGS sequence, were then separated from the remaining His_6-tagged SUMO and His_6-tagged human rhinovirus 3C protease by re-loading onto the 2 × 5 mL HisTrap HP columns pre-equilibrated in buffer A. Purification was then carried out with the following program: sample loading, 1 mL min^{-1}; washing, 2 mL min^{-1}, 4 column volumes of buffer A; elution, 2 mL min^{-1}, 2 column volumes of buffer B; elution, 2 mL min^{-1}, 2 column volumes of buffer C. The untagged docking domains passed through the column during the washing step, and were collected and concentrated to 5−7 mL using an Amicon Ultra 3000 MWCO centrifuge filter (Millipore Corp).

Subsequently, the concentrated docking domains were loaded onto a size exclusion chromatography column (Superdex 75 26/60 column, GE Healthcare) equilibrated in buffer GF (20 mM HEPES, 100 mM NaCl, 0.5 mM TCEP, pH 7.5). Following a concentration step, the purity of the purified proteins was verified by SDS-PAGE, and their concentrations were determined by NanoDrop (Thermo Scientific) with extinction coefficients calculated using the ExPASy ProtParam tool[63].

Isothermal Titration Calorimetry measurements. ITC measurements were performed at 20 °C in buffer GF using a MicroCal ITC200 (Malvern Instruments) (A2F Plateforme ASIA: Approches fonctionnelles et Structurales des InterActions cellulaires). A 300 μL aliquot of $^{N}DD_5$ at 70 μM was placed in the calorimeter cell and titrated with 700 μM of the $^{C}DD_4$s ($^{C}DD_4$ wild type, $^{C}DD_4$ SDM and $^{C}DD_4$ helix swap) in the syringe. In the case of the binding experiments between $^{N}DD_9$ Met and $^{C}DD_8$, the $^{C}DD_8$ (700 μM) was added to the $^{N}DD_9$ Met (80 μM in the cell), while for the binding between $^{N}DD_9$ Val and the ^{C}DDs ($^{C}DD_8$, $^{C}DD_4$ wild type, $^{C}DD_4$ SDM and $^{C}DD_4$ helix swap), the ^{C}DDs (700 μM) were added to $^{N}DD_9$ Val in the cell (120 μM). The ITC experiments were then carried out as followed: initial waiting time 120 s, initial injection of 0.5 μL over 1 s followed by 19 serial injections of 2 μL over 4 s, separated by an interval of 120 s. For each experiment, the reference power was set to 5 μcal^{-1}, stirring speed to 750 rpm, and the high feedback mode was selected. Two independent titrations were performed for each combination of DDs. The heat of reaction per injection ($\mu cal\ s^{-1}$) was determined by integration of the peak areas using the Origin 7.0 (OriginLab) software, assuming a one-site binding model (consistent with the solved structures of the types of DDs[22,27]), yielding the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N) and the dissociation constant (K_d). The heats of dilution of the DDs were determined by injecting them into the cell containing buffer only, and these were subtracted from the corresponding binding data prior to curve fitting.

In some cases, when a plateau (binding saturation) was not reached at the final titration step, and the problem could not be solved by increasing the concentration of DD in syringe, we initially placed $^{C}DD/^{N}DD$ complex in the ITC cell (at the concentration of the two partners reached in the previous titration), filled the syringe with additional DD, and performed a second titration experiment. This procedure was then repeated until binding saturation was reached. To fit the data, the MicroCal ITC concatenation software was used to combine two ITC data files together. Most importantly, the critical parameter dimensionless constant (C-value) was calculated as follows:

$$C = NK_a[M]_T$$

where K_a is the binding constant, $[M]_T$ is the total macromolecular concentration in the cell, and N is the stoichiometry of interaction. A reliable ITC binding isotherm is evidenced by ITC data with C-values >1 (the optimal range is $5 < C < 500$)[64], as was the case for all of our measurements.

Circular Dichroism measurements. CD spectra were recorded on a Chirascan CD (Applied Photophysics, United Kingdom) (IBS-Lor UMS 2008 Plateforme de Biophysique et Biologie Structurale) at 0.5 nm intervals in the wavelength range of 180−260 nm at 20 °C, using a temperature-controlled chamber. A 0.01 cm quartz cuvette containing 30 μL of docking domain at 100 μM, a 0.1 cm cuvette with 200 μL of sample at 10 μM, and a 1 cm cuvette containing 1.5 mL of sample at 1 μM, were used for all the measurements. All measurements were performed at least in triplicate, and sample spectra were corrected for buffer

background by subtracting the average spectrum of buffer alone. The CD spectra were deconvoluted using the deconvolution software CDNN2.1[65] to estimate the secondary structure present in the docking domains.

SEC-MALS analysis of docking domains. The oligomeric state of all the docking domains was determined by size-exclusion chromatography multi-angle light scattering (SEC-MALS) on the A2F Plateforme ASIA. For this, SEC was first carried out on a Superdex75 10/300 column (GE Healthcare) at 20 °C using a flow rate of 0.5 mL min^{-1} in HEPES buffer (20 mM HEPES, 100 mM NaCl, 0.5mM TCEP, pH 7.5) using a ÄKTA-Purifier FPLC (GE Healthcare). Multi-angle light scattering (MALS) was measured using a MiniDAWN TREOS II (Wyatt Technology), while refractometry was monitored using an Optilab T-rEX (Wyatt Technology). Data processing was carried out with the manufacturer supplied software (ASTRA 6.1, Wyatt Technology) to determine the protein oligomerization state.

HPLC-MS analysis of fermentation metabolites and purified docking domains. The fermentation broth of *Streptomyces* was centrifuged at 4000*g* for 10 min. As described previously[23], the stambomycins and their derivatives were then extracted from the mycelia, by first resuspending the cells in 40 mL distilled water, followed by centrifugation (4000*g*, 10 min, repeated 3×) to remove water-soluble components. After decanting the water, the cell pellets were weighed and extracted with methanol by shaking at 150 rpm for 2 h at room temperature. Thereafter, the methanol extracts were filtered to remove the cell debris, followed by rotary evaporation to dryness. The obtained extracts were then dissolved in methanol, whose volume was determined according to the initial weight of the mycelia (70 μL methanol to 1 g of initial cell pellet). The resulting mycelial crude extracts were then passed through a 0.4 μm syringe filter and analyzed in positive electrospray mode (ESI^+) by HPLC-HRMS (Thermo Scientific Orbitrap LTQ XL or an Orbitrap ID-X Tribrid Mass Spectrometer) (Plateau d'Analyse Structurale et Métabolomique (PASM)-SF4242 EFABA) using an Alltima™ C18 column (2.1 × 150mm, 5 μm particle size). Separation was carried out with Milli-Q water containing 0.1% formic acid (A) and acetonitrile containing 0.1% formic acid (B) using the following elution profile: 0−48 min, linear gradient 5−95% solvent B; 48−54 min, constant 95% solvent B; 54−60 min, constant 5% solvent B. Mass spectrometry operating parameters were: spray voltage, 5 kV; source gases were set respectively for sheath gas, auxiliary gas and sweep gas at 30, 10 and 10 arbitrary units min^{-1}; capillary temperature, 275 °C; capillary voltage, 4 V; tube lens, split lens and front lens voltages 155 V, −28 V and −6 V, respectively. Due to the much lower sensitivity of the Orbitrap LTQ XL relative to the Orbitrap ID-X Tribrid as evidenced by comparative analysis of identical samples on the two instruments, we introduced a 10× correction factor to the the yields determined using the Orbitrap LTQ XL (**Supplementary Table 5**).

The purified docking domains in buffer GF were diluted with Milli-Q water to a concentration of 50 μM and injected onto an Alltima™ C18 column (2.1 × 150mm, 5 μm particle size). Analysis was carried out with Milli-Q water containing 0.1% TFA (A) and acetonitrile containing 0.1% TFA (B), using the elution profile:

0−15 min, constant 10% solvent B; 15−20 min, linear gradient of 10% solvent B to 95%; 20−25 min, constant 10% solvent B. Mass spectrometry operating parameters were set as above.

Metabolite profiling of engineered strains. Comparative analysis of fermentation extracts (of all strains containing pOE484 except K7N3, CPN4 and CPN5, relative to control mutant containing empty plasmid pIB139) was conducted on a Dionex Ultimate 3000 HPLC system coupled to a Thermo Scientific™ Q Exactive™ Hybrid Quadrupole-Orbitrap mass spectrometer. MS-settings: spray voltage 3.5 kV; capillary temperature 320 °C; sheath gas (52.50), auxiliary gas (13.75), sweep gas (2.75); probe heater 437.50 °C; S-Lens RF (50), positive mode, resolution 70.000; AGC target 1e6, microscans 1, maximum IT 75 ms, scan range 200−1800 *m*/*z*. Chromatographic separation was obtained using a Phenomenex Kinetex 2.6 µm XB-C18 150 × 4.6 mm column with solvents (A, H_2O + 0.1% formic acid) and (B, MeCN + 0.1% formic acid) and the following gradient: flow rate 0.7 mL min^{-1}, 20% B for 2 min, 20% to 98% B over 18 min, 98% B for 5 min, 98% to 20% B in 0.5 min and 20% B for 4 min. Metabolic differences within the obtained data (**Supplementary Table 6**) were identified using SIEVE 2.0 screening software, applying the default settings for component extraction of small molecules, except that of the base peak minimum intensity, which was set to 5000000.

Quantification of metabolites. To quantify the yields of native stambomycins and the newly-generated derivatives by HPLC-MS, we generated a calibration curve with previously-purified stambomycin A/B **1** as the standard (using a concentration range between 1 µg L^{-1} and 50 mg L^{-1}). This approach yielded a linear correlation between the quantity of **1** and the respective peak area in the extracted ion chromatogram (EIC) (the areas of all peaks corresponding to the parental ions were used) (**Supplementary Fig. 12**, **Supplementary Table 5**). The titers of all stambomycin derivatives were then determined using this calibration curve, based on the peak areas corresponding to the parental ions in their respective EIC chromatograms (**Supplementary Table 5**). As the limit of detection in these experiments lay between 1 and 10 µg L^{-1}, yields in this range must be viewed as estimates.

Acknowledgements

We acknowledge financial support from the Agence Nationale de la Recherche (grant number ANR-16-CE92-0006-01, PKS STRUCTURE to K.J.W.), the Université de Lorraine, the Centre National de la Recherche Scientifique (CNRS), and the IMPACT Biomolecules project of the Lorraine Université d'Excellence (Investissements d'avenir – ANR 15-004). J.P. acknowledges funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742739). S. Collin and B. Chagot are thanked for help with the analysis of the DDs *in vitro*, A. Kriznik for

assistance with the biophysical experiments, J. Grosjean for help with the HPLC-MS analysis and S. Rousselot for contributing to developing the genetic engineering strategy.

Author contributions

L.S. constructed plasmids, generated and fermented recombinant strains, analyzed and interpreted HPLC-MS data, expressed and purified recombinant docking domains, carried out biophysical analysis of DD interactions, and generated all of the manuscript figures. L.H. and C.J. constructed certain plasmids, mutant BACs and/or mutant strains. C.P. performed HPLC analyses. A.B. and J.P. carried out HPLC-MS analyses and metabolic profiling. B.A., C.J. and K.J.W. designed the research and supervised the project. K.J.W. performed *in silico* analyses and wrote the manuscript, with contributions from all authors.

Additional information

Supplementary Information accompanies this paper at **Appendix to chapter II.**

References

1. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670 (2015).
2. Keatinge-Clay, A. T. The uncommon enzymology of *cis*-acyltransferase assembly lines. *Chem. Rev.* **117**, 5334–5366 (2017).
3. Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* 33, 203–230 (2016).
4. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
5. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).
6. Hans, M., Hornung, A., Dziarnowski, A., Cane, D. E. & Khosla, C. Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules. *J. Am. Chem. Soc.* **125**, 5366–5374 (2003).
7. Ranganathan, A. *et al.* Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues. *Chem. Biol.* **6**, 731–741 (1999).
8. McDaniel, R., Kao, C. M., Hwang, S. J. & Khosla, C. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4**, 667–674 (1997).
9. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).
10. Peng, H., Ishida, K., Sugimoto, Y., Jenke-Kodama, H. & Hertweck, C. Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases. *Nat. Commun.* **10**, 3918 (2019).
11. Miyazawa, T., Hirsch, M., Zhang, Z. & Keatinge-Clay, A. T. An *in vitro* platform for engineering and harnessing modular polyketide synthases. *Nat. Commun.* **11**, 80 (2020).
12. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
13. Wu, N., Tsuji, S. Y., Cane, D. E. & Khosla, C. Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* **123**, 6465–6474 (2001).
14. Zhang, L. *et al.* Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. *Angew. Chem. Int. Ed Engl.* **56**, 1740–1745 (2017).
15. Keatinge-Clay, A. T. Polyketide synthase modules redefined. *Angew. Chem. Int. Ed Engl.* **56**, 4658–4660 (2017).
16. Chandran, S. S., Menzella, H. G., Carney, J. R. & Santi, D. V. Activating hybrid modular interfaces in synthetic polyketide synthases by cassette replacement of ketosynthase domains. *Chem. Biol.* **13**, 469–474 (2006).
17. Cortes, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487–1489 (1995).
18. Massicard, J.-M., Soligot, C., Weissman, K. J. & Jacob, C. Manipulating polyketide stereochemistry by exchange of polyketide synthase modules. *Chem. Commun.* **56**, 12749–12752 (2020).

19. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22066–22071 (2010).
20. Klaus, M. *et al.* Protein-protein Interactions, not substrate recognition, dominate the turnover of chimeric assembly line polyketide synthases. *J. Biol. Chem.* **291**, 16404–16415 (2016).
21. Klaus, M., Buyachuihan, L. & Grininger, M. The ketosynthase domain constrains the design of polyketide synthases. *ACS Chem. Biol.* **15**, 2422–2432 (2020).
22. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
23. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6258–6263 (2011).
24. Aigle, B., Challis, G., Laureti, L., Song, L. & Leblond, P. Stambomycin and derivatives, their production and their use as drugs. Patent WO2011009938A2 (2011).
25. Wang, S., Dong, G. & Sheng, C. Structural simplification of natural products. *Chem. Rev.* **119**, 4180–4220 (2019).
26. Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4110–4115 (2012).
27. Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
28. Nguyen, T. *et al.* Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
29. Murphy, A. C., Hong, H., Vance, S., Broadhurst, R. W. & Leadlay, P. F. Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem. Commun.* **52**, 8373–8376 (2016).
30. Weissman, K. J. The structural basis for docking in modular polyketide biosynthesis. *ChemBioChem* **7**, 485–494 (2006).
31. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1541–1546 (2003).
32. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723–728 (2015).
33. Wilkinson, C. J. *et al.* Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417–426 (2002).
34. Song, L. *et al.* Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis. *J. Antibiot.* **67**, 71–76 (2014).
35. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207–220 (2001).
36. Kittendorf, J. D., Beck, B. J., Buchholz, T. J., Seufert, W. & Sherman, D. H. Interrogating the molecular basis for multiple macrolactone ring formation by the pikromycin polyketide synthase. *Chem. Biol.* **14**, 944–954 (2007).
37. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402−W407 (2019).
38. Risser, F. *et al.* Towards improved understanding of intersubunit interactions in modular polyketide biosynthesis: Docking in the enacyloxin IIa polyketide synthase. *J. Struct. Biol.* 212, 107581 (2020).
39. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
40. Dorival, J. *et al.* Characterization of intersubunit communication in the virginiamycin *trans*-acyl transferase polyketide synthase. *J. Am. Chem. Soc.* **138**, 4155–4167 (2016).
41. Porterfield, W. B., Poenateetai, N. & Zhang, W. Engineered biosynthesis of alkyne-tagged polyketides by type I PKSs. *iScience* **23**, 100938 (2020).
42. Yuzawa, S. *et al.* Comprehensive *in vitro* analysis of acyltransferase domain exchanges in modular polyketide synthases and its application for short-chain ketone production. *ACS Synth. Biol.* **6**, 139–147 (2017).
43. Wlodek, A. *et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Commun.* **8**, 1206 (2017).
44. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin.* **47**, 231–243 (2015).
45. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4**, 1020–1029 (2015).
46. Zeng, H. *et al.* Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA(sm) combined system. *Appl. Microbiol. Biotechnol.* **99**, 10575–10585 (2015).

47. Zhang, J. J. & Moore, B. S. Site-directed mutagenesis of large biosynthetic gene clusters via oligonucleotide recombineering and CRISPR/Cas9 targeting. *ACS Synth. Biol.* **9**, 1917–1922 (2020).
48. Myronovskyi, M., Rosenkränzer, B. & Luzhetskyy, A. Iterative marker excision system. *Appl. Microbiol. Biotechnol.* **98**, 4557–4570 (2014).
49. Mo, J. *et al.* Efficient editing DNA regions with high sequence identity in actinomycetal genomes by a CRISPR-Cas9 system. *Synth. Syst. Biotechnol.* **4**, 86–91 (2019).
50. Tong, Y. *et al.* Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20366–20375 (2019).
51. Zhao, Y., Li, G., Chen, Y. & Lu, Y. Challenges and advances in genome editing technologies in *Streptomyces*. *Biomolecules* **10**, (2020).
52. Alberti, F. & Corre, C. Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat. Prod. Rep.* **36**, 1237–1248 (2019).
53. Wang, Q. *et al.* Dual-function chromogenic screening-based CRISPR/Cas9 genome editing system for actinomycetes. *Appl. Microbiol. Biotechnol.* **104**, 225–239 (2020).
54. Ye, S., Enghiad, B., Zhao, H. & Takano, E. Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in *Streptomyces*. *J. Ind. Microbiol. Biotechnol.* **47**, 413–423 (2020).
55. Bozhüyük, K. A. J. *et al.* Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661 (2019).
56. Aparicio, J. F. *et al.* Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**, 9–16 (1996).
57. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
58. Sambrook, J., Fritsch, E. F. & Maniatis, T. Molecular cloning: a laboratory manual. (Cold Spring Harbor Laboratory Press, New York, 1989).
59. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, Norwich, 2000).
60. Raynal, A., Karray, F., Tuphile, K., Darbon-Rongère, E. & Pernodet, J.-L. Excisable cassettes: new tools for functional analysis of *Streptomyces* genomes. *Appl. Environ. Microbiol.* **72**, 4839–4844 (2006).
61. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).
62. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68 (1992).
63. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).
64. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137 (1989).
65. Böhm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5**, 191–195 (1992).
66. Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. 14, 898–908 (2007).

Chapter III Harnessing the inherent AT promiscuity of the stambomycin PKS for the synthesis of analogues

3.1 Introduction

Acyl transferase (AT) domains of type I modular PKSs select specific extender units for incorporation into polyketide scaffolds, and thus govern to a large extent the branching functionality within polyketide structures. They thus offer a powerful potential means for the regioselective diversification of polyketide skeletons, especially in the case of ATs which exhibit atypical substrate specificity (e.g. AT_4 of FK506 PKS (specific for allylmalonyl-CoA), AntD-AT of antimycin PKS (specific for butylmalonyl-CoA and hexylmalonyl-CoA) [168, 169]. To date, a variety of approaches including AT domain swapping, *trans*-AT complementation, and active-site mutagenesis have been applied to polyketide structural engineering with varying levels of success [63] (additional examples are described in **Chapter I, section 1.7.1**).

Experimentally, AT swapping consists of exchanging only the AT domain within a module of interest for an AT of different substrate specificity from another module (either sourced from the same or a different PKS) (**Figure 3.1a**). This approach has resulted in various hybrid modules, with recent work notably claiming to have identified "optimal' sites for such exchanges [127]. *trans*-AT complementation involves inactivating the native AT within a target module (AT^0 shown in **Figure 3.1b**) and complementing it with a *trans*-acting AT. However, this approach requires that the *trans*-AT tolerate the noncognate ACP substrate, otherwise the complementation will be at best inefficient [102]. Finally, AT active site mutagenesis is used to try to alter or extend the innate AT substrate selectivity (**Figure 3.1c**). Despite the availability of several AT crystal structures [106, 143, 170, 171] this approach remains hampered by the paucity of co-complex structures in the presence of native extender units, and indeed, all attempts so far to switch AT specificity in this way have only resulted in a shift in substrate preference despite in some cases large numbers of introduced mutations [105].

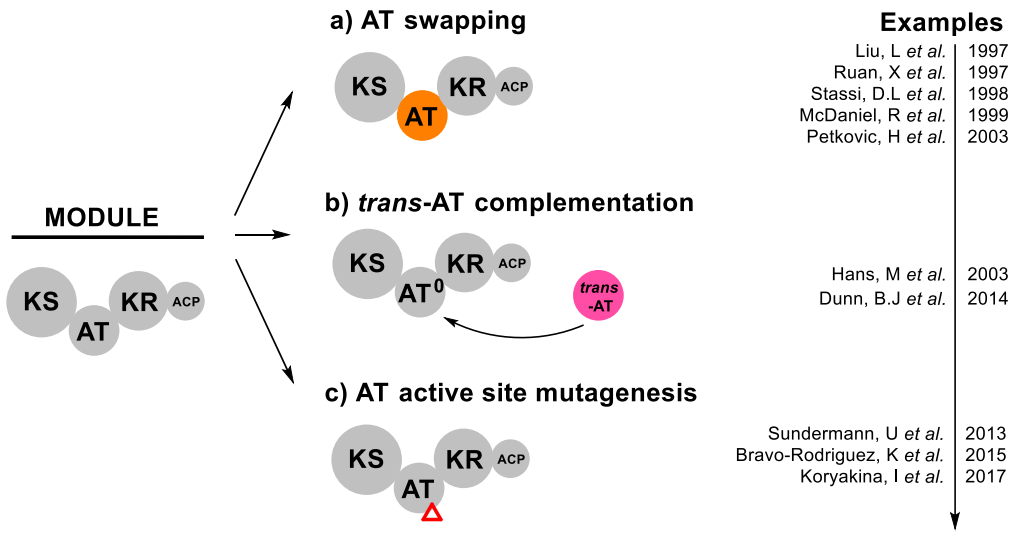


Figure 3.1 Strategies for PKS engineering by altering AT domain substrate specificity. Examples cited in this figure are [97-102, 105, 172-174]

A second way to exploit AT domains towards polyketide structural diversification is to exploit the intrinsically broad substrate specificity of certain ATs, and challenge them with alternative building blocks via precursor-directed biosynthesis (PDB) and/or mutasynthesis (this point has been developed in **section 1.7.6**). To date, multiple functional groups including several not found naturally in polyketides have been incorporated into polyketides in this way, including alkyne, azide, and halogens moieties [148-151, 175-178].

The critical prerequisite for applying PDB and/or mutasynthesis to polyketide structural engineering is a supply of exogenous building blocks. Previous *in vitro* studies have shown that commercially available malonyl-CoA derivatives can be used to achieve incorporation. However, this method is not suitable for *in vivo* feeding studies due to the lack of acyl-CoA membrane permeability. To solve the problem, synthetic *N*-acetylcysteamine thioesters (SNAC, mimics of CoA) are often used for feeding experiments, but leveraging them in large-scale applications is complicated by the need to synthesize the compounds, as well as their cell toxicity [144]. A promising alternative is a chemoenzymatic approach (see **Chapter I, section 6.6**), in which a fed monomer (mono- or diacid) is activated to its corresponding malonyl-CoA derivative *in cellulo* by an appropriate enzyme (either a crotonyl-CoA reductase/carboxylase (CCR) homologue or a malonyl-CoA synthetase (MatB) homologue (**Figure 3.2**).

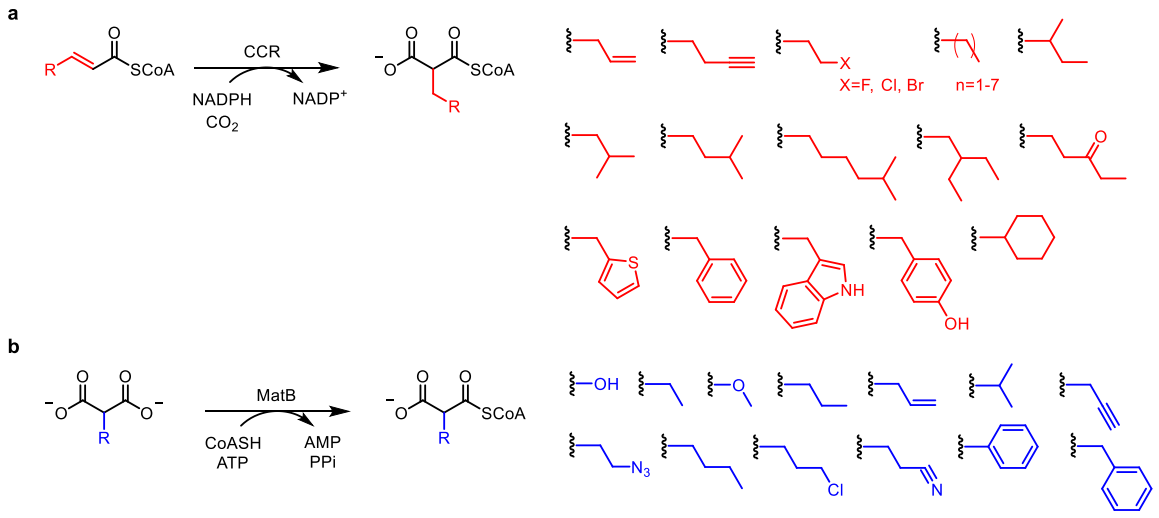


Figure 3.2 Alternative extender unit generation for polyketide engineering *via* the crotonyl-CoA reductase/carboxylase (CCR) and MatB machineries. a) CCR enzymes (e.g. AntE, CinF, SalG and SpnE, etc.) and their engineered variants catalyze the reductive carboxylation of CoA linked α, β-unsaturated acyl-CoA via utilization of one reducing equivalent of NADPH, to afford malonyl-CoA analogous. **b**) MatB (e.g. MatB_Sc from *S. coelicolor*, MatB_cinna from *S. cinnamonesis* and MatB_Rt from *Rhizobium trifolii*, etc.) and derived mutants are capable of generating the CoA derivatives of a broad range of malonate analogues, via an ATP-dependent reaction.

As described in **section 1.8**, the biosynthesis of stambomycins A–F, a family of 51-membered macrolides in *S. ambofaciens* ATCC23877 [40], involves an exceptional AT domain, AT_{12}, which recruits at least six atypical alkyl malonyl-CoA-based extender units. This broad specificity results in six different stambomycins bearing different side chains at C-26, a position which has been shown to directly impact the antiproliferative properties of the metabolites [40]. Recently, the Challis group described an unprecedented pathway for extender unit generation operating in stambomycin biosynthesis [179], which comprises two ATP-dependent enzymes, SamR0482 and SamR0483 (**Figure 3.3**). SamR0482 first activates the medium chain acid to its corresponding CoA, which is then directly carboxylated to yield the malonyl-CoA analogue. Both SamR0482/SamR0483 exhibit substrate tolerance, which notably allowed incorporation via PDB of azide- and alkyne-containing extender units into the stambomycins via the broad specificity AT_{12}, as well as improvement in yields of the natural structure stambomycin E by feeding of heptanoic acid (**Figure 3.3**) (under native conditions, the yields of stambomycins E and F are much lower than A–D, **Appendix Figure 3.1**). Notably analogues bearing both azide

and alkyne groups are attractive targets for further structural elaboration by 'click chemistry', providing a promising progress for further action study of stambomycin.

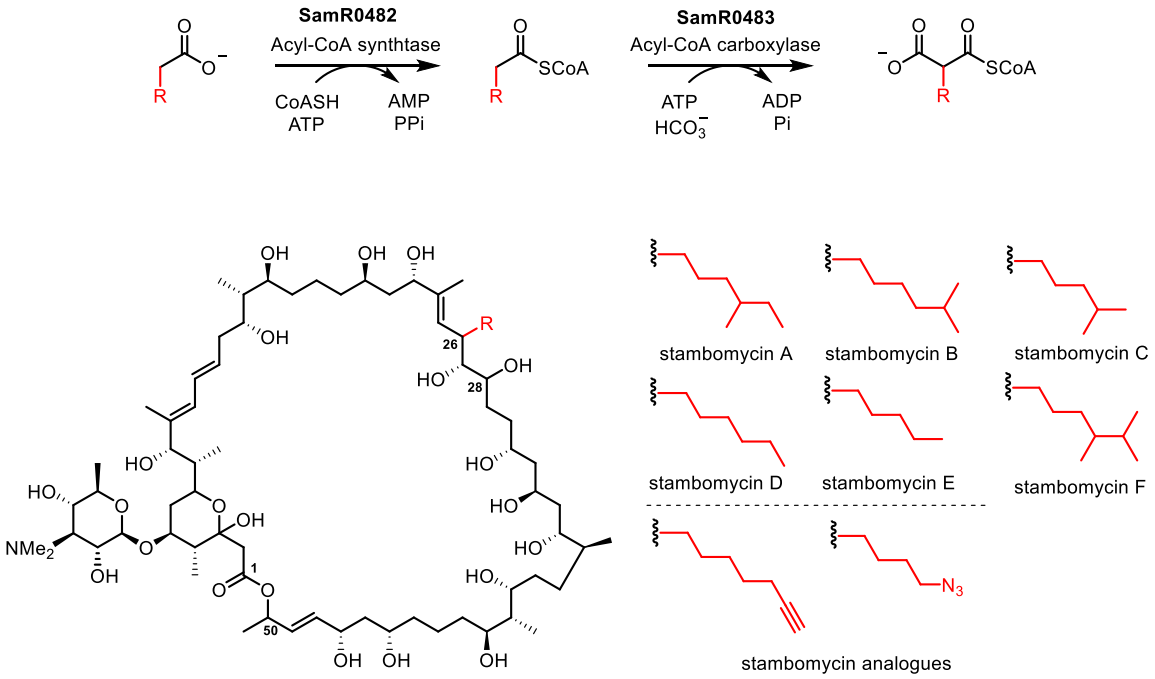


Figure 3.3 The SamR0482/SamR0483 mediated pathway, and the structures of native and PDB-derived novel stambomycins. The two-enzyme mediated pathway affords six atypical malonyl-CoA extender units, by CoA derivatization followed by direct carboxylation of medium chain acyl-carboxylic acids sourced from the cellular pool. It can also act on exogenously-fed carboxylic acids, giving rise to the corresponding stambomycin analogues.

As an alternative, we aimed to establish stambomycin mutasynthesis as a means to access additional analogues for biological evaluation, but in addition to potentially improve yields relative to the previous PDB experiment by removing competition from the native extender units. For this, we aimed to disable the endogenous pathway to the six unusual extender units by inactivating SamR0482 and/or SamR0483, again relying on AT_{12} for incorporation of exogenous malonyl-CoA derivatives. Encouragingly, BLAST analysis of the *S. ambofaciens* genome sequence, revealed one enzyme (called MatB_Sa) with an amino acid identity of 90%, 75% and 39% respectively with MatB_Sc (*S. coelicolor*) [83], MatB_cinna (*S. cinnamonesis*) [144], and MatB_Rt (*Rhizobium trifolii*) [153]. This finding suggested that MatB_Sa might also possess the necessary broad specificity to efficiently activate non-native malonate derivatives to their corresponding CoAs, but we also carried

out experiments in which the bonafide promiscuous MatB_cinna (kind gift of F Schultz, Ruhr-University, Bochum, Germany) was introduced into the engineered mutasynthesis strain.

3.2 Results

3.2.1 Construction of mutants ATCC/OE484/Δ482 and ATCC/OE484/Δ483

In order to disable the native pathway to the six extender units, the genes encoding SamR0482 and SamR0483 were deleted using PCR-targeting (**section 3.4.2**), resulting in the mutants ATCC/Δ482 (constructed by Laureti [180]) and ATCC/Δ483. The identity of mutant ATCC/Δ483 in which the gene was replaced by an 81 bp scar was confirmed by PCR and DNA sequencing (**Figure 3.4**). Subsequently, the LAL regulator overexpressing plasmid pOE484 was introduced into the mutants as previously described [40], affording the mutants ATCC/OE484/Δ482 (constructed by Luisa LAURETI) and ATCC/OE484/Δ483 (this work) (**Table 3.2**).

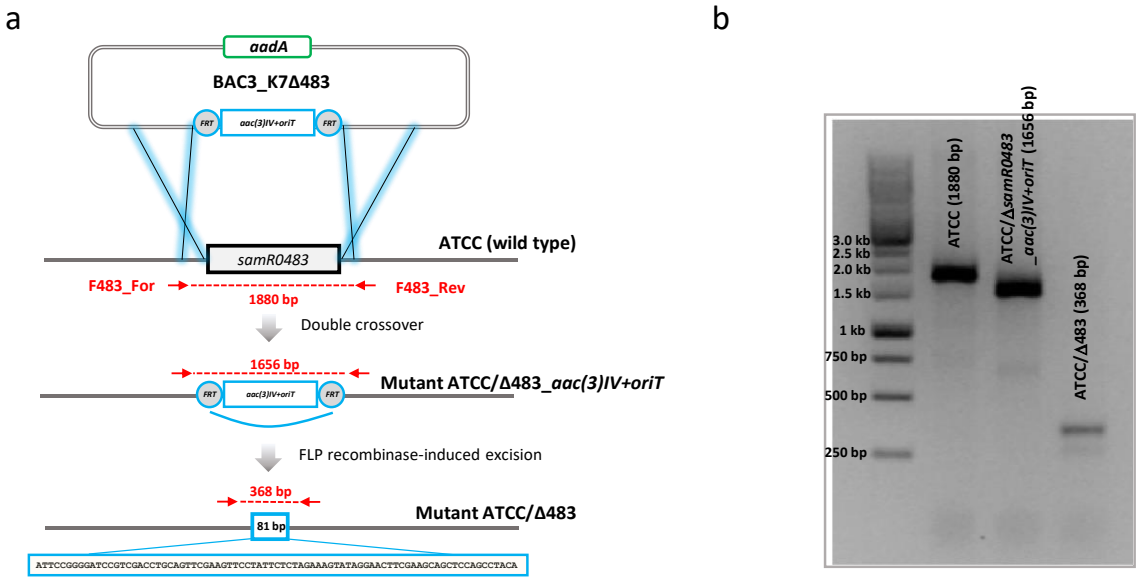


Figure 3.4 In-frame deletion of *samR0483* in *S. ambofaciens* using PCR-targeting genetic manipulation, and the downstream PCR screening/verification strategy. a) Schematic representation of double crossover based on homologous recombination. The horizontal arrows indicate primer locations which were used to identify the wild type and double crossover progeny after cassette excision. **b**) Agarose gel showing the PCR amplicons using flanking primers (F484_For/F483_Rev) annealing to the chromosome upstream

and downstream of the target gene in the wild type genome, cassette pre-excised progenies and cassette excised progenies, respectively.

3.2.2 HPLC-MS analysis of mutants ATCC/OE484/Δ482 and ATCC/OE484/Δ483

The mutants ATCC/OE484/Δ482 and ATCC/OE484/Δ483 were cultivated in MP5 for 4 days at 30 °C and then the mycelia were harvested and extracted with methanol. The extracts of mutants ATCC/OE484/Δ482 and ATCC/OE484/Δ483 were analyzed by HPLC-MS in comparison to the control strain ATCC/OE484. Both the mass and UV (λ_{max} 238 nm) chromatograms show that production of the native stambomycins in mutant ATCC/OE484/Δ483 has been almost completely abolished (**Figure 3.5**), consistent with successful disruption of the precursor supply to AT_{12}. On the other hand, mutant ATCC/OE484/Δ482 still behaved essentially like the wild type strain, although the yield of stambomycins was reduced to 59%. BlastP analysis of the *S. ambofaciens* genome using SamR0482 as a query revealed several homologues, two of which are located in the antimycin and congocidine clusters, respectively: AntF (acyl-CoA synthetase, 29.8% amino acid identity to SamR0482) and Cgc22 (acyl-CoA synthetase, 34.5% amino acid identity to SamR0482). We thus propose that the loss-of-function of SamR0482 can be compensated for by either AntF, Cgc22 or both (**Table 3.1**). Similarly, the BlastP analysis of SamR0483 also revealed several homologues, especially the duplicate AlpX (proposed to be involved in kinamycin biosynthesis) and two PccBs (propionyl-CoA carboxylase, PCC β-subunit) which show high mutual amino acid sequence identity (**Table 3.1**). Typically, PccBs act on short chain acyl-CoAs (**section 1.6.2.1.1**), potentially explaining why only a low level of SamR0483 complementation was observed (0.4 % stambomycins relative to wild type).

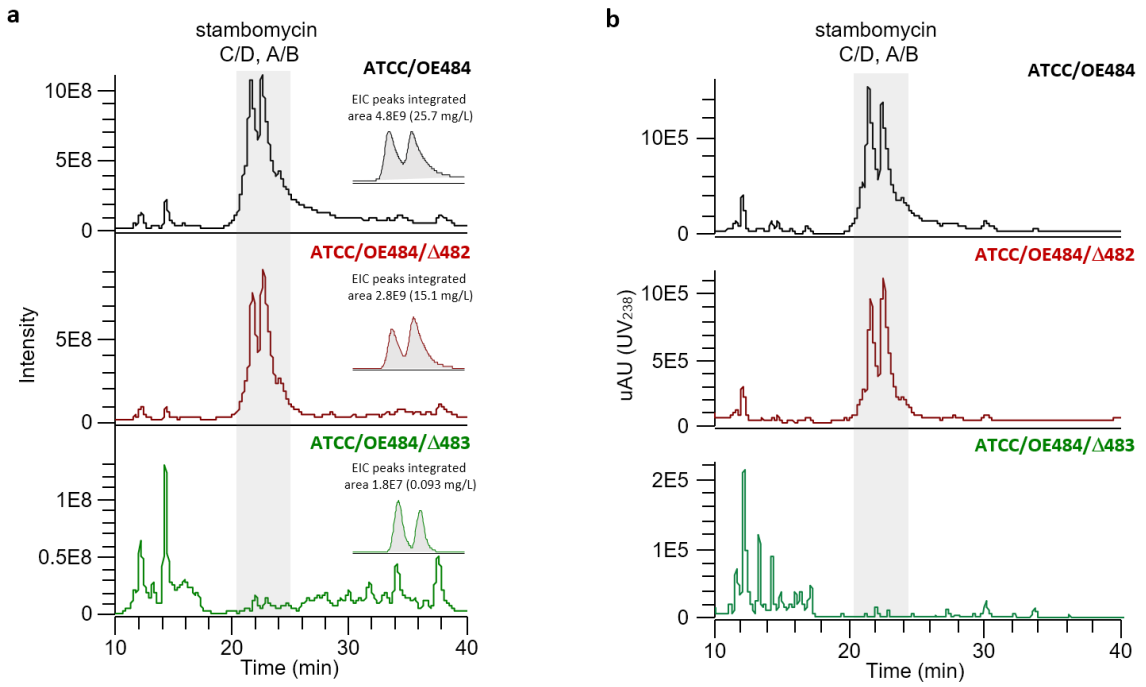


Figure 3.5 Analysis by (a) mass spectrometry and (b) UV at λ_{max} 238 nm of strain ATCC/OE484 (control) and mutants ATCC/OE484/Δ482, and ATCC/OE484/Δ483. The mass chromatogram and UV spectrum of ATCC/OE484/Δ483 clearly show that production of the stambomycins (peaks with retention times of 20−24 min, highlighted in grey) is abolished. On the basis of integration of the EIC peaks (the calculated area shown here is the mean of the analysis of at least two independent clones of each strain), the relative yield of stambomycins in mutant ATCC/OE484/Δ482 is approximately 59% of that of the wild type, while only 0.4% was detected from mutant ATCC/OE484/Δ483.

Table 3.1 Homologues of SamR0482 and SamR0483 in *S. ambofaciens* ATCC23877 revealed by BlastP analysis

Identities	Genome coordinates	Features	Accession number
Homologs of SamR0482 in *S. ambofaciens* ATCC23877			
29.8%	SAM23877_0423	AntF (acyl-CoA ligase, involved in antimycin biosynthesis pathway)	AKZ53472.1
34.5%	SAM23877_6654	Cgc22 (acyl-CoA synthetase, involved in congocidine biosynthesis pathway)	AKZ59699.1
27.8%	SAM23877_2750	FadD (Long-chain-fatty-acid-CoA ligase)	AKZ55799.1
28.1%	SAM23877_3675	AlkK (Medium-chain-fatty-acid-CoA ligase)	AKZ56720.1
Homologs of SamR0483 in *S. ambofaciens* ATCC23877			
52%	SAM23877_0214 and SAM23877_7459	AlpX (putative carboxyl transferase, involved in kinamycin biosynthesis)	AKZ53263.1

59%	SAM23877_4747	PccB (propionyl-CoA Carboxylase β chain)	AKZ57792.1
52%	SAM23877_5290	PccB (propionyl-CoA Carboxylase β chain)	AKZ58335.1
31%	SAM23877_2811	Propionyl-CoA carboxylase	AKZ55860.1
29%	SAM23877_3346	Methyl crotonyl-CoA carboxylase carboxyl transferase	AKZ56393.1
26%	SAM23877_2503	Acetyl-coenzyme A carboxylase carboxyl transferase	AKZ55552.1

3.2.3 HPLC-MS analysis of mutant ATCC/OE484/Δ483::*samR0483*

To further confirm that the substantial decrease in stambomycin production in the ATCC/OE484/Δ483 mutant was directly due to the loss function of SamR0483, we conducted a complementation experiment (**Figure 3.13**). The *samR0483* gene was PCR amplified and cloned under the control of the strong, constitutive *ermEp** promoter in the conjugative and integrative vector pOSV809 (**Table 3.3**). The resulting construct pOSV809_PermE_SamR0483 was integrated into the genome of the ATCC/OE484/Δ483 mutant, giving rise to ATCC/OE484/Δ483::*samR0483* (**Figure 3.13**). The newly-generated mutant was subjected to the same procedure for fermentation and HPLC-MS analysis as described earlier. This analysis demonstrated that the complemented mutant produced 70-fold more stambomycins than the ATCC/OE484/Δ483 mutant, but that complementation was nonetheless incomplete, as the complemented strain produced only 35% of the wild type yields. Taken together with the previous *in vitro* investigation of SamR0482 and SamR0483 [179], these data proved that the *samR0483* gene is essential for stambomycin biosynthesis, and that the in-frame deleted mutant ATCC/OE484/Δ483 was a good base strain for carrying out stambomycin mutasynthesis (**Figure 3.6**).

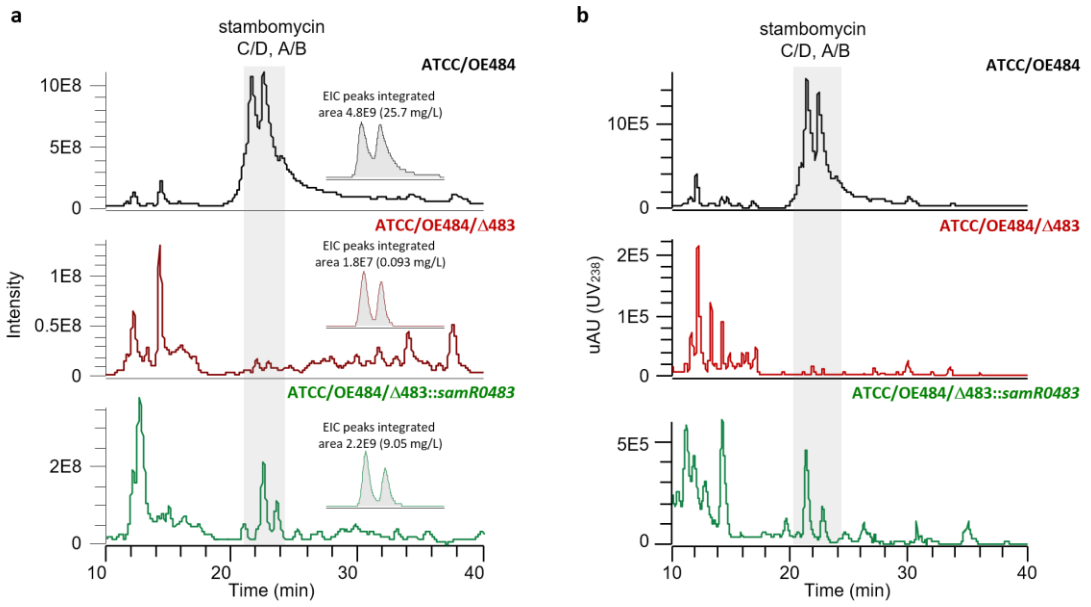


Figure 3.6 Analysis by (a) mass spectrometry and (b) UV at λ_{max} 238 nm of strain ATCC/OE484, ATCC/OE484/Δ483 and the complementation mutant ATCC/OE484/Δ483::*samR0483*. The chromatograms resulting from analysis of ATCC/OE484/Δ483::*samR0483* clearly show that production of the stambomycins was restored (peaks with retention times of 22–24 min, highlighted in grey (the minor discrepancy in retention time is due to peak shape divergence induced by the large difference in yields. The mass spectral data are identical for the corresponding peaks (data not shown). On the basis of integration of the EIC peaks (the calculated area shown here is the mean of at least two independent clones of each strain), the relative yield of stambomycins in the complementation mutant ATCC/OE484/Δ483::*samR0483* is approximately 35% of the wild type production levels.

3.2.4 Mutasynthesis of stambomycin analogues

Separate cultures of 50 mL MP5 medium in 250 mL Erlenmeyer flasks were inoculated from a single preculture of the ATCC/OE484/Δ483 mutant strain and cultivated at 30 °C, 200 rpm for 24 h. Allylmalonic acid (**Figure 3.7**) was then added to the culture broth at two different concentrations (5 mM and 10 mM), following by an additional 3 days of cultivation. Subsequently, extracts were obtained from mycelia and analyzed by HPLC-MS using the same procedure as described earlier, in parallel with a control (*i.e.* a culture of the mutant strain without any added allylmalonic acid). For verification purposes, each feeding experiment was conducted at least in triplicate. In this case, analysis of the obtained data revealed ions in the allyl-fed extracts with *m*/*z* = 1318.8606 and 1302.8650, in addition to those corresponding to stambomycins CD and AB (*m*/*z* =1362.9227 and 1376.9392

[M+H]$^+$). These new ions correspond to the calculated mass for allyl-stambomycin ([M+H]$^+$: $C_{69}H_{124}NO_{22}^+$, Calc'd: 1318.8610, Mass error = −0.30 ppm, minute amounts detected) and allyl-deoxystambomycin ([M+H]$^+$: $C_{69}H_{124}NO_{21}^+$, Calc'd: 1302.8660, Mass error = −0.76 ppm, main product), an intermediate of stambomycin without the hydroxylation catalyzed by the P450 hydroxylase SamR0478 at C28 [163]) (**Figure 3.7** and **3.8**). In comparison with the fermentation extract of the ATCC/OE484/Δ483 mutant grown without feeding, the yields of the parental stambomycins from the 5 mM fed extract were approximately 5-fold higher (0.6% relative to the wild type to 3%, as judged on the basis of the integrated area of EIC (extracted ion chromatogram) peaks). To explain this surprising observation, we propose that the fed allylmalonic acid somehow induces expression of the SamR0483 homologues, which can then more substantially compensate for the loss of SamR0483. Intriguingly, when the concentration of fed allylmalonic acid was increased to 10 mM, the yields of stambomycins were reduced compared to the 5 mM fed extract. In addition, the production of allyl-deoxystambomycin improved by 2.3-fold relative to the 5 mM fed extract. This finding suggests a preferred feeding dosage of 10 mM to facilitate incorporation of exogenous extender unit.

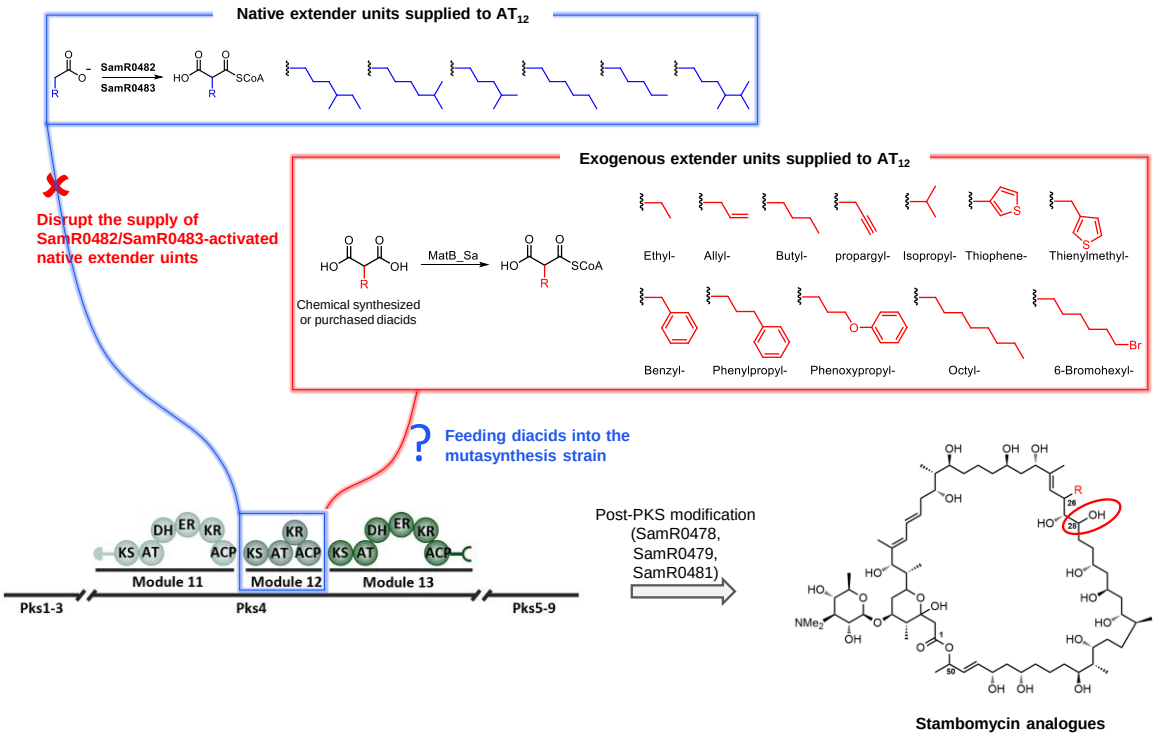
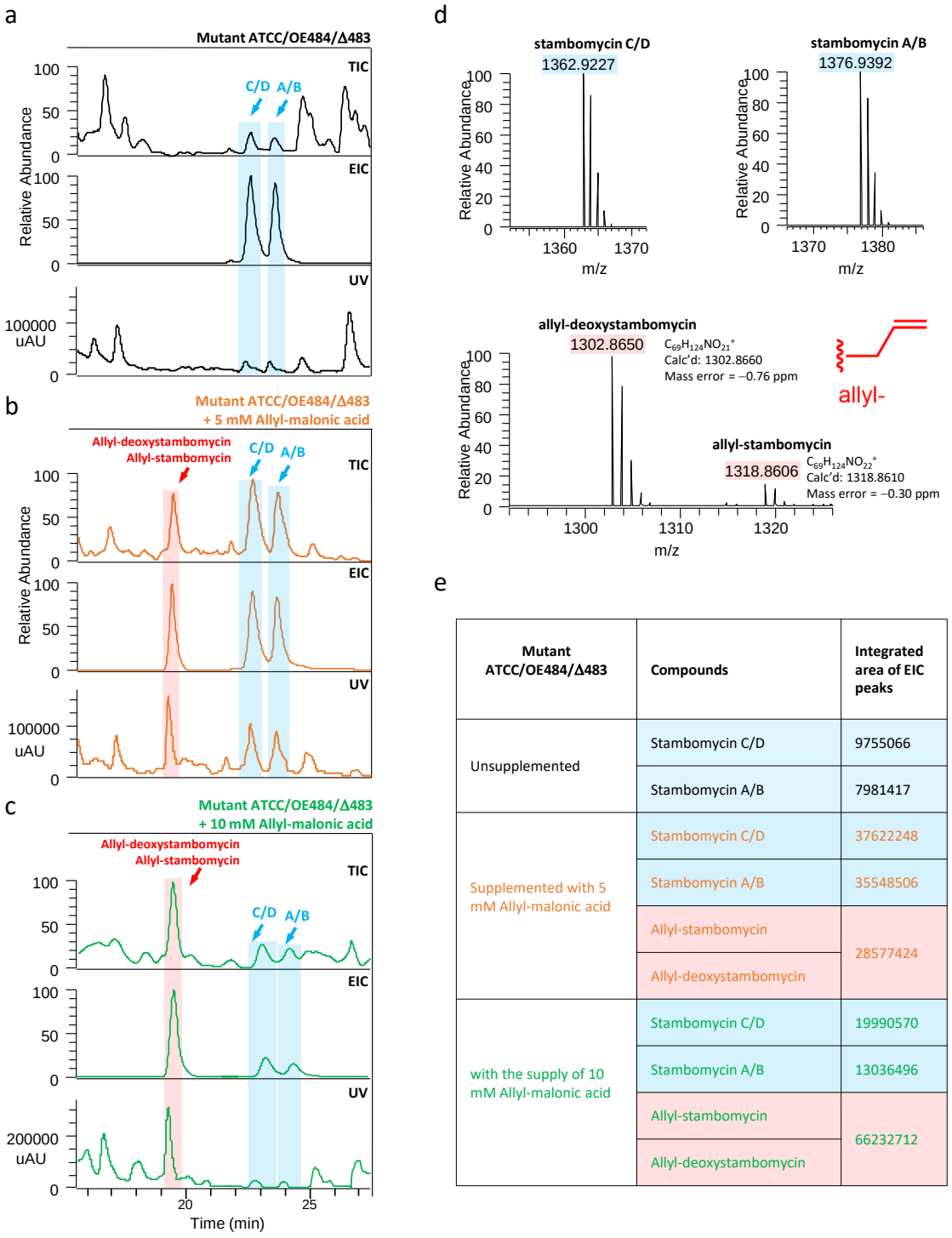


Figure 3.7 Schematic overview of mutasynthesis. Malonic acids incorporating a wide variety of side chains are fed to the fermentation medium of mutasynthesis mutant ATCC/OE484/Δ483. If the acids are successfully activated to their corresponding CoAs by the promiscuous MatB_Sa, they can become substrates for the broad specificity AT_{12} domain. Successful incorporation by AT_{12} results in stambomycins modified at C-26. Note, the post-PKS hydroxylation at C-28 (red circled) occurs to variable extents.



e

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Unsupplemented	Stambomycin C/D	9755066
	Stambomycin A/B	7981417
Supplemented with 5 mM Allyl-malonic acid	Stambomycin C/D	37622248
	Stambomycin A/B	35548506
	Allyl-stambomycin	28577424
	Allyl-deoxystambomycin	
with the supply of 10 mM Allyl-malonic acid	Stambomycin C/D	19990570
	Stambomycin A/B	13036496
	Allyl-stambomycin	66232712
	Allyl-deoxystambomycin	

Figure 3.8 Analysis by mass spectrometry and UV at λ_{max} 238 nm of crude extract from the ATCC/OE484/Δ483 mutant and mutant fed with allylmalonic acid. a) TIC (total ion chromatogram), EIC (extracted ion chromatogram) and UV spectrum of the control without supplementary allylmalonic acid. **b**) TIC, EIC and UV spectra of the mutant ATCC/OE484/Δ483 **s**upplied with allylmalonic acid at a final concentration of 5 mM. **c**) TIC, EIC and UV spectra of the mutant ATCC/OE484/Δ483 supplied with allylmalonic acid at a final concentration of 10 mM. **d**) Mass spectra of stambomycins C/D, A/B, allyl-deoxystambomycin and allyl-stambomycin. **e**) Comparison of the yields of stambomycins and the allyl-incorporated stambomycin in the three extracts based on the integrated area of the EIC peaks. For all integrated areas, the mean value from at least 3 fermentation results is given in the table.

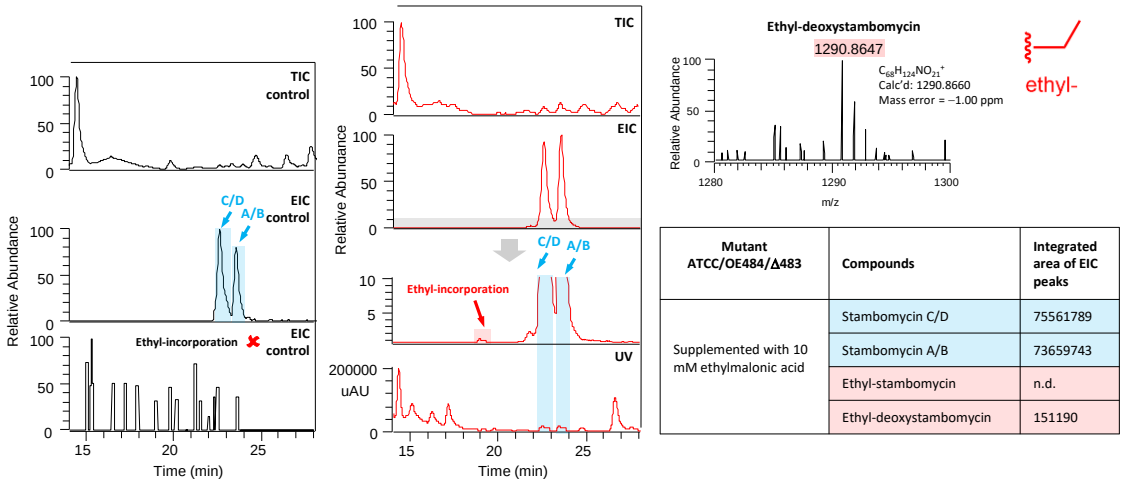
In theory, the incorporation of alternative extender units depends on three distinct factors: the ability of MatB_Sa to activate them as their CoA thioesters, the tolerance of AT_{12} towards the modified structures, and whether the altered chain extension intermediates are accepted by all downstream modules. The successful incorporation of allyl-malonic acid provided initial evidence that these three checkpoints exhibit at least some level of tolerance towards non-native substrates. Indeed, this result was anticipated for the PKS components, as AT_{12} and the downstream modules recognize six different native chains.

Following on from this experiment, cultures of the ATCC/OE484/Δ483 mutant strain were supplemented with a wider panel of diacids, including ethyl-, isopropyl-, butyl-, benzyl-, octyl-, phenylpropyl-, phenoxypropyl-, thiophene-, thienylmethyl-, propargyl-, and 6-bromohexyl-malonic acids, at a final concentration of 10 mM. The analytical data showed that the supplementation with ethyl-, butyl-, benzyl-, phenoxypropyl-, and 6-bromohexyl-malonic acids yielded new peaks relative to the control, all with masses consistent with the expected ethyl-, butyl-, benzyl-, phenoxypropyl, and 6-bromohexyl-stambomycin analogues (**Figure 3.9a**, **c**, **d**, **g, k**). In contrast, the phenylpropyl-, thiophene-, thienylmethyl-, and propargyl-malonic acids failed to be incorporated (**Figure 3.9b, f, h, i, j**). Furthermore, a mass ion of 1390.9546 ($[M+H]^+$, $C_{74}H_{130}NO_{22}^+$) found in the octyl-fed extract (**Figure 3.9e**) requires further investigation, in order to determine whether it corresponds to a novel octyl-stambomycin or the native stambomycin F [161] (the two masses are identical). Among the five successful incorporations (ethyl-, allyl-, butyl-, benzyl- and phenoxypropyl-), the stambomycin analogues lacking the C-28 hydroxylation (the 'deoxy' forms) were the main or even sole products (i.e. no ethyl-stambomycin was

detected), typically accounting for 80−90% of the total yields based on their relative abundance in the mass spectra. In contrast, 6-bromohexyl-stambomycin is the sole analogue produced in the 6-bromohexyl feeding experiments. This observation suggests that the P450 SamR0478 strongly prefers its native substrates or at minimum closely-similar chains (**Figure 3.7**). In addition, in no case were aglycone analogues detected in the fed extracts (data no shown), showing that the post-PKS glycosylase SamR0481, exhibits broad substrate tolerance (**Figure 3.7**).

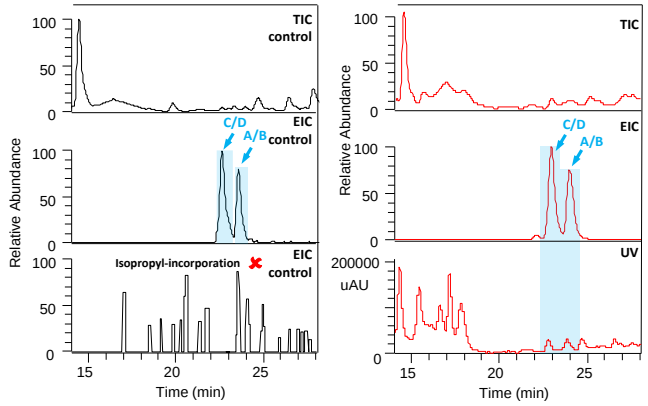
Relative to the minute production of ethyl-, benzyl-, phenoxypropyl-, and 6-bromohexyl-incorporated stambomycin analogues, the allyl-deoxystambomycin and butyl-deoxystambomycin shown strong UV absorbance at 238 nm (**Figure 3.8c and Figure 3.9c**), suggesting that purification and structural characterization of these metabolites may be possible. For all fed-extracts with the exception of thiophene-malonic acid and thienylmethyl-malonic acid for which there was no change, the masses corresponding to stambomycins A/B and C/D were always detected, and at yields 2–7 times higher than from the unsupplemented ATCC/OE484/Δ483 mutant.

a



Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM ethylmalonic acid	Stambomycin C/D	75561789
	Stambomycin A/B	73659743
	Ethyl-stambomycin	n.d.
	Ethyl-deoxystambomycin	151190

b

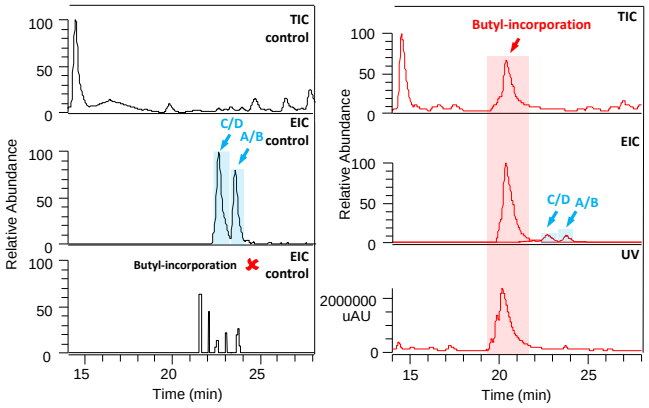


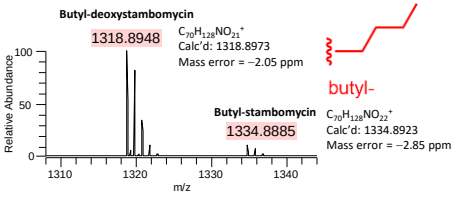
No detectable isopropyl-incorporation

isopropyl-

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM isopropyl-malonic acid	Stambomycin C/D	36206957
	Stambomycin A/B	27020195
	Isopropyl-stambomycin	n.d.
	Isopropyl-deoxystambomycin	

c

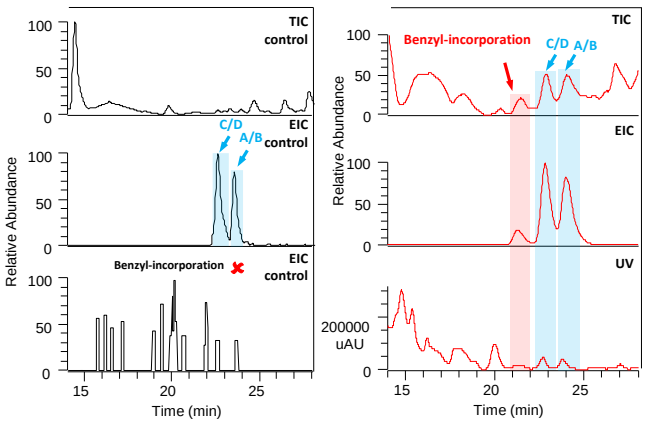


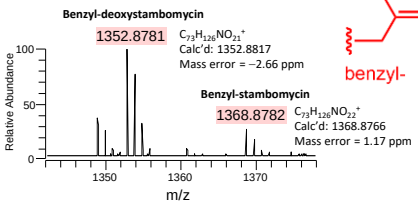


butyl-

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM butyl-malonic acid	Stambomycin C/D	25164939
	Stambomycin A/B	8403620
	Butyl-stambomycin	757196925
	Butyl-deoxystambomycin	3608738285

d

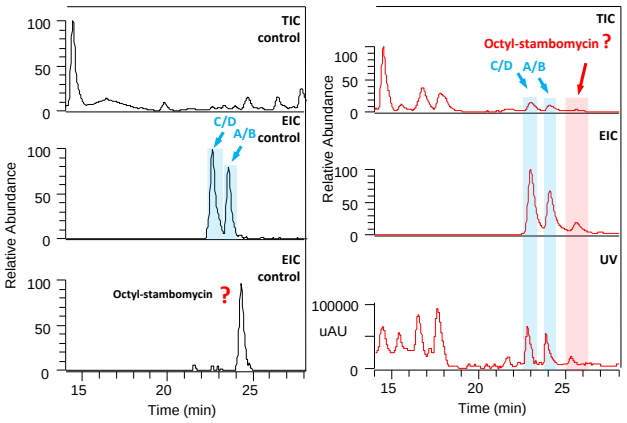


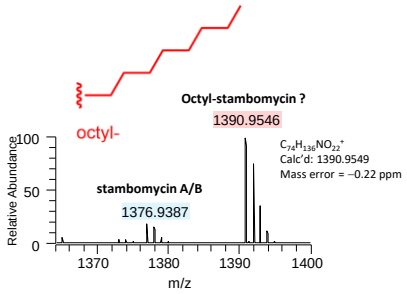


benzyl-

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of UV peaks
Supplemented with 10 mM benzyl-malonic acid	Stambomycin C/D	43247606
	Stambomycin A/B	40353586
	Benzyl-stambomycin	4814150
	Benzyl-deoxystambomycin	9595426

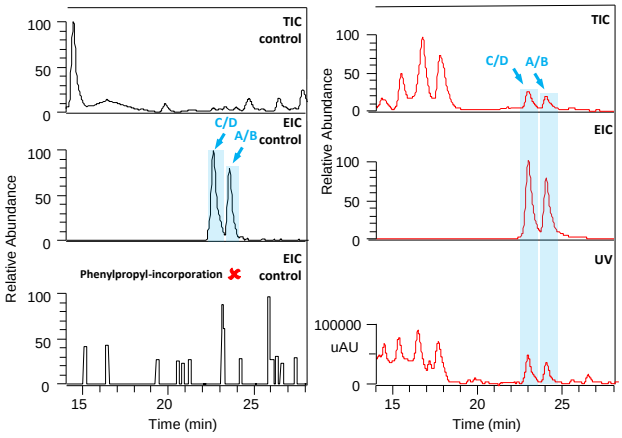
e





Mutant ATCC/OE484/Δ483	Compounds	Integrated area of UV peaks
Supplemented with 10 mM octyl-malonic acid	Stambomycin C/D	89591504
	Stambomycin A/B	58322496
	Octyl-stambomycin	13572281
	Octyl-deoxystambomycin	n.d.

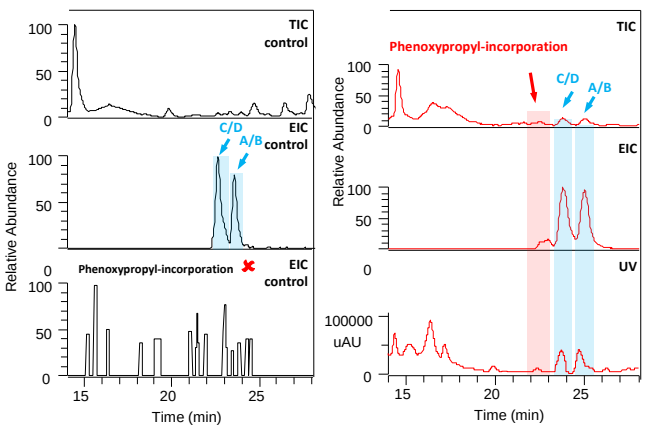
f

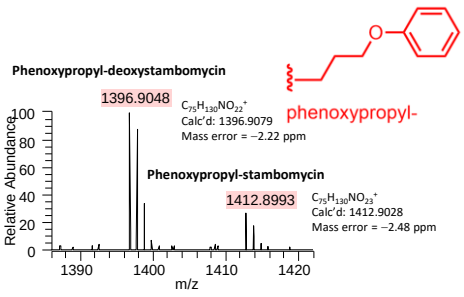


No detectable phenylpropyl-incorporation

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of UV peaks
Supplemented with 10 mM phenylpropyl-malonic acid	Stambomycin C/D	60638995
	Stambomycin A/B	42790287
	Phenylpropyl-stambomycin	n.d.
	Phenylpropyl-deoxystambomycin	

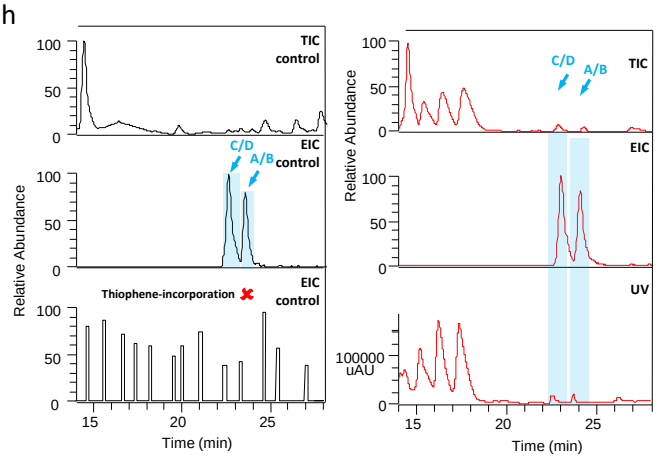
g





Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM phenoxypropyl-malonic acid	Stambomycin C/D	21692929
	Stambomycin A/B	19007029
	Phenoxypropyl-stambomycin	1916650
	Phenoxypropyl-deoxystambomycin	2366663

h

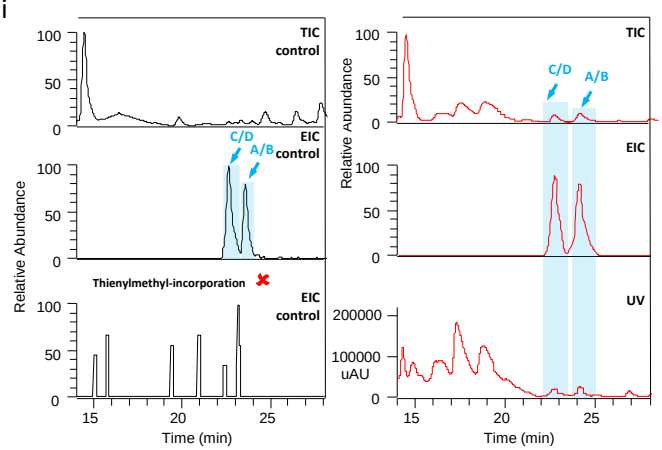


thiophene-

No detectable thiophene-incorporation

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM thiophene-malonic acid	Stambomycin C/D	7480600
	Stambomycin A/B	7211573
	Thiophene-stambomycin	n.d.
	Thiophene-deoxystambomycin	

i

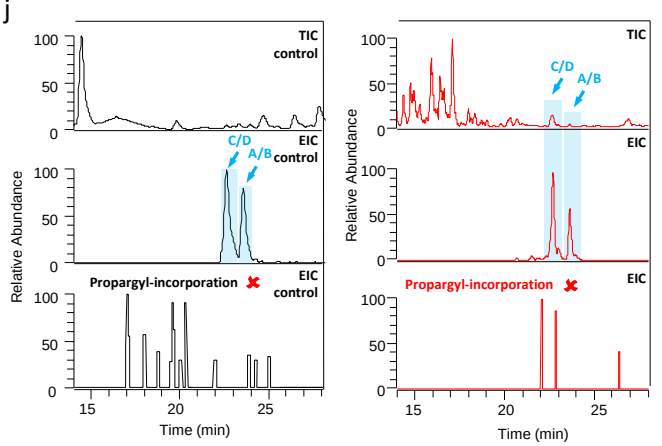


thienylmethyl-

No detectable thienylmethyl-incorporation

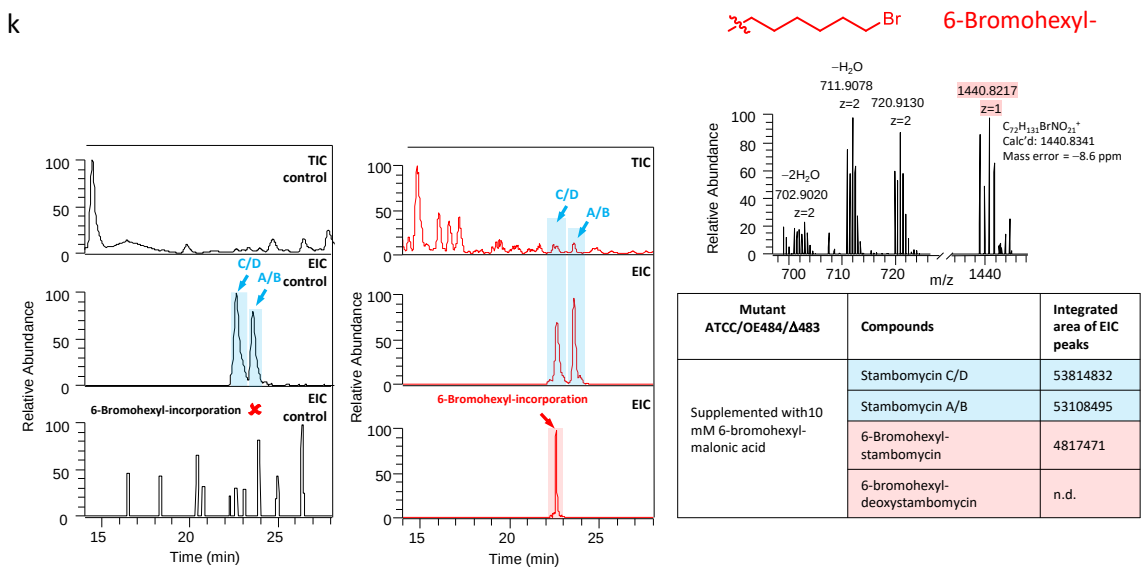
Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM thienylmethyl-malonic acid	Stambomycin C/D	9347943
	Stambomycin A/B	8056985
	Thienylmethyl-stambomycin	n.d.
	Thienylmethyl-deoxystambomycin	

j



propargyl-

Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM 6-bromohexyl-malonic acid	Stambomycin C/D	53595673
	Stambomycin A/B	26483512
	Propargyl-stambomycin	n.d.
	Propargyl-deoxystambomycin	n.d.



Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with10 mM 6-bromohexyl-malonic acid	Stambomycin C/D	53814832
	Stambomycin A/B	53108495
	6-Bromohexyl-stambomycin	4817471
	6-bromohexyl-deoxystambomycin	n.d.

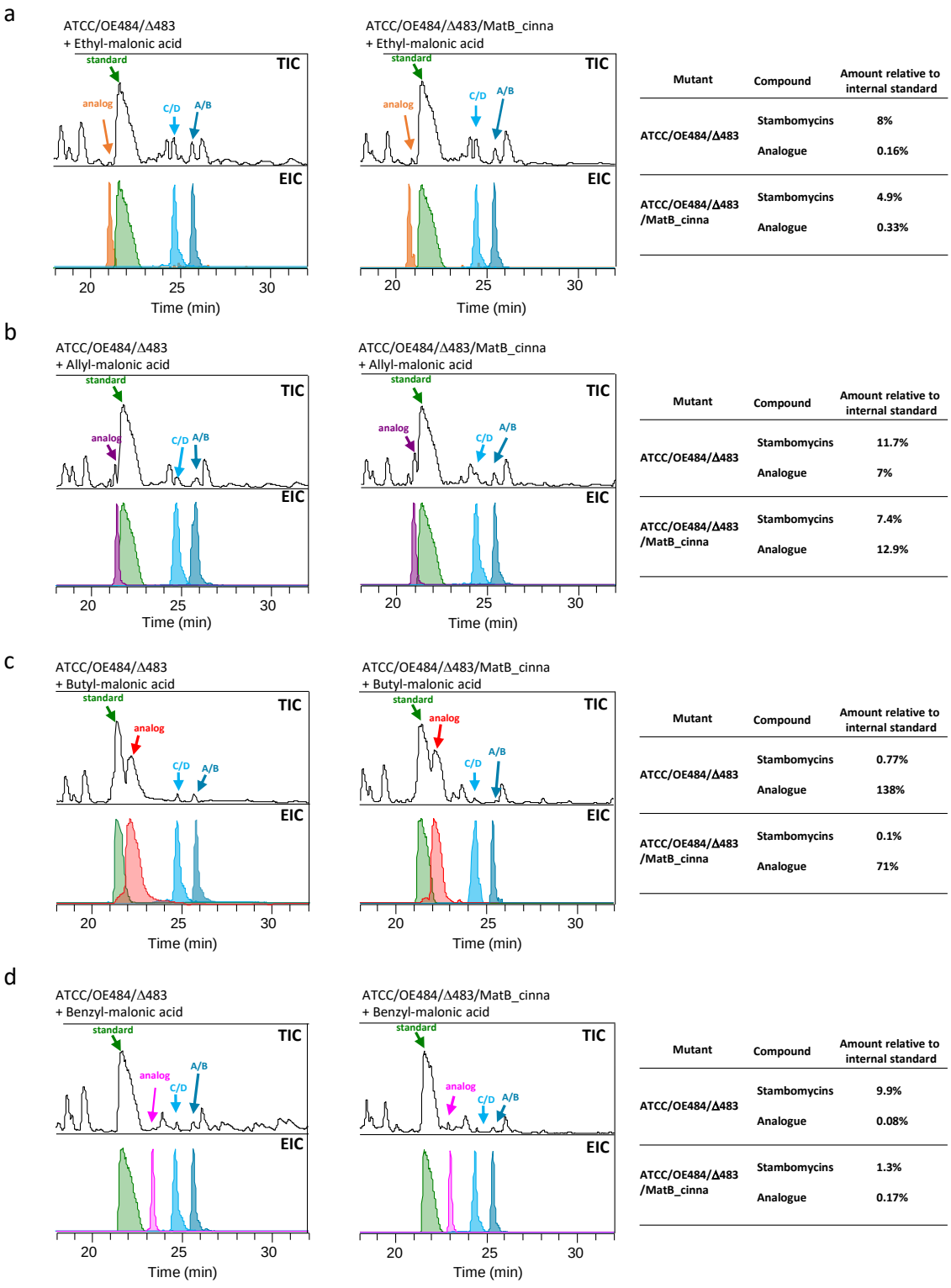
Figure 3.9 Analysis by mass spectrometry and UV detection of crude extract from the mutasynthesis strain ATCC/OE484/Δ483 supplemented with various malonic acid alternatives in comparison to the control without supplementation. In each case are shown: the TIC, EIC and UV_{238} of crude fed-extracts, the mass spectrum of the corresponding stambomycin analogues, and the production levels of both the analogues and the native stambomycins A–D based on the integrated area of EIC peaks (average of two independent fermentations). "n.d." indicates compounds which were not detected.

3.2.5 Quantitative analysis of the stambomycin derivatives produced by the mutasynthesis mutants ATCC/OE484/Δ483 and ATCC/OE484/Δ483/MatB_cinna

When considered that the lack of production of mature stambomycin derivatives in the presence of isopropyl-, phenyl-, thiophene-, thienylmethyl-, and propargyl-malonic acid, might have resulted from intolerance to the modified chains by downstream modules. As such a mechanism could have resulted in premature release of stalled intermediates, we analyzed for the presence in the extracts of intermediates corresponding to the products of modules 12 and 13. However, no convincing signal was detected. Furthermore, we did not detect any of the predicted shunt products in extracts from feeding experiments where the stambomycin analogues were obtained, consistent with the idea that the modules downstream of module 12 are broadly tolerant to alternative chains. In light of this evidence, we hypothesize that the major blocks to incorporation of diacids are failure to efficiently activate the fed substrates to their CoA forms by MatB_Sa, or of the AT_{12} to recognize the modified malonates. To directly address the first possibility, we introduced into mutant

ATCC/OE484/Δ483 the plasmid pRT801_lacZ_PermE_MatB_cinna (a kind gift of Prof. F Schulz, Ruhr-Universität Bochum) encoding the broad specificity MatB from Streptomyces *cinnomonensis*, resulting in mutant ATCC/OE484/Δ483/MatB_cinna. Indeed, recent work by Frank's group [144] has demonstrated that MatB_cinna could efficiently convert ethyl-, allyl- and benzyl-malonic acid to their corresponding CoA forms. We reasoned, therefore, that the new mutant could be a better starting point for feeding experiments compared to ATCC/OE484/Δ483.

Mutants ATCC/OE484/Δ483 and ATCC/OE484/Δ483/MatB_cinna were then fed ethyl-, allyl-, butyl- and benzyl-malonic acids, followed by quantification of the resulting analogues produced by the two strains. For this, erythromycin (1 mM added during the work-up step) was used as an internal standard for the HPLC-MS analyses to control for any differences in extraction efficiency (**section 3.4.4**). The obtained data demonstrated that the presence of MatB_cinna improved the incorporation of ethyl-, allyl- and benzyl-malonic acids (**Figure 3.10**), resulting in a nearly 2-fold increase relative to the parental strain (i.e. that containing only MatB_Sa). Incorporation of butyl-malonic acid by the two mutants essentially identical, reflecting the fact that MatB_cinna does not activate butyl-diacid.



Mutant	Compound	Amount relative to internal standard
ATCC/OE484/Δ483	Stambomycins	8%
	Analogue	0.16%
ATCC/OE484/Δ483 /MatB_cinna	Stambomycins	4.9%
	Analogue	0.33%

Mutant	Compound	Amount relative to internal standard
ATCC/OE484/Δ483	Stambomycins	11.7%
	Analogue	7%
ATCC/OE484/Δ483 /MatB_cinna	Stambomycins	7.4%
	Analogue	12.9%

Mutant	Compound	Amount relative to internal standard
ATCC/OE484/Δ483	Stambomycins	0.77%
	Analogue	138%
ATCC/OE484/Δ483 /MatB_cinna	Stambomycins	0.1%
	Analogue	71%

Mutant	Compound	Amount relative to internal standard
ATCC/OE484/Δ483	Stambomycins	9.9%
	Analogue	0.08%
ATCC/OE484/Δ483 /MatB_cinna	Stambomycins	1.3%
	Analogue	0.17%

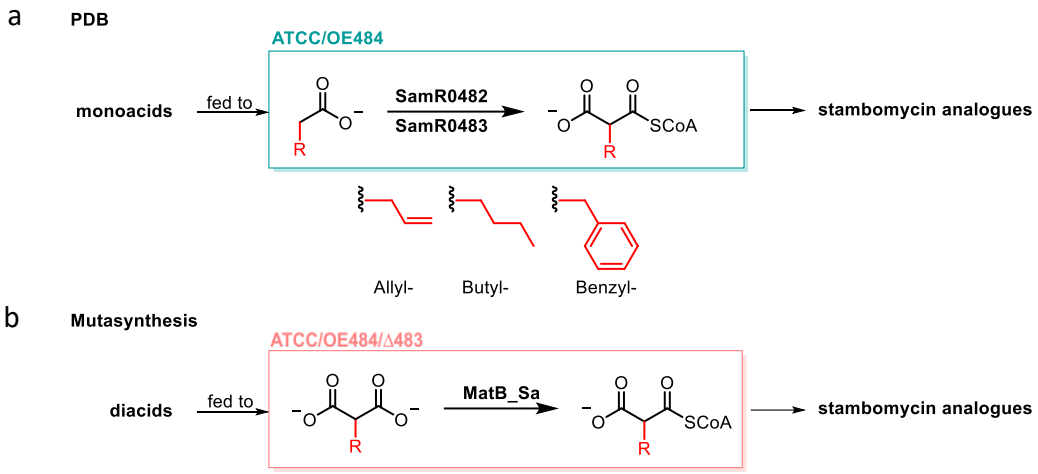
Figure 3.10 Quantitative analysis of the mutasynthesis mutants ATCC/OE484/Δ483 and ATCC/OE484/ Δ483/MatB_cinna by HPLC-MS using erythromycin as the internal standard. a) Comparison of the two mutants when supplied with ethyl-malonic acid. **b–d**) Analysis following

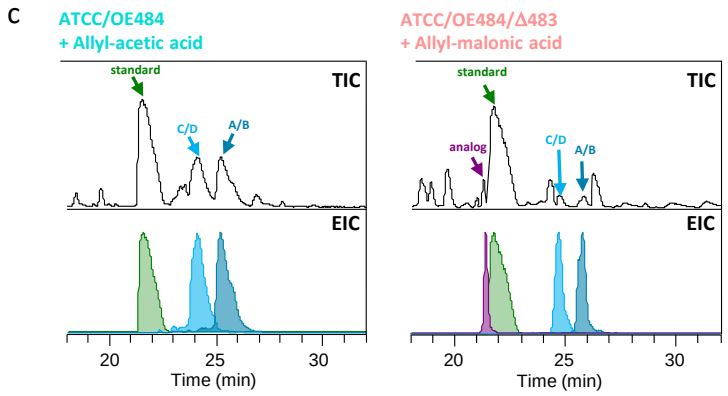
supplementation with allyl-, butyl- and benzyl-malonic acids. TIC (total ion chromatogram) of the crude extract is shown in the top panel in each case. The amount of stambomycins (C/D+A/B) and stambomycins analogues relative to the internal standard (Ery) were determined by the integrated area of EIC peaks (mean value of three individual fermentation results) and the peak aera of standard was taken as 100%.

3.2.6 Quantitative comparative analysis of the PDB and mutasynthesis approaches by HPLC-MS, using erythromycin as an internal standard

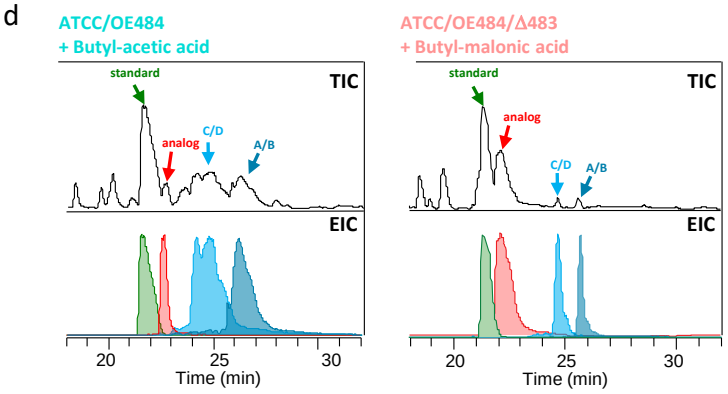
As we observed that the parental stambomycins continued to be produced by the mutasynthesis strain, it was of interest to compare the relative efficiency of the approach to the alternative PDB strategy. In contrast to mutasynthesis, PDB requires fed mono-acids, which are then activated by the SamR0482/SamR0483 pair to generate the corresponding extender unit CoA thioesters. To directly compare the two pathways, we compared feeding of strain ATCC/OE484 with allyl-, butyl, and benzyl-acetic acids (**Figure 3.11a**), and mutant ATCC/OE484/Δ483 with allyl-, butyl- and benzyl-malonic acids (**Figure 3.11b**). Erythromycin (1 mM) was again used as internal standard in the HPLC-MS analyses to permit accurate quantification.

The obtained data showed that butyl-acetic acid was activated by the SamR0482/SamR0483-mediated pathway in strain ATCC/OE484, resulting in butyl-incorporated stambomycin analogues (**Figure 3.11d**), but that incorporation of allyl- and benzyl-acetic acid failed (**Figure 3.11c, e**). Moreover, the yield of butyl-incorporated analogues from the ATCC/OE484/Δ483 mutant relative to ATCC/OE484 (7.7E10 vs 7.4E9, **Figure 3.11d**), confirmed that the mutasynthesis was indeed the more efficient approach.

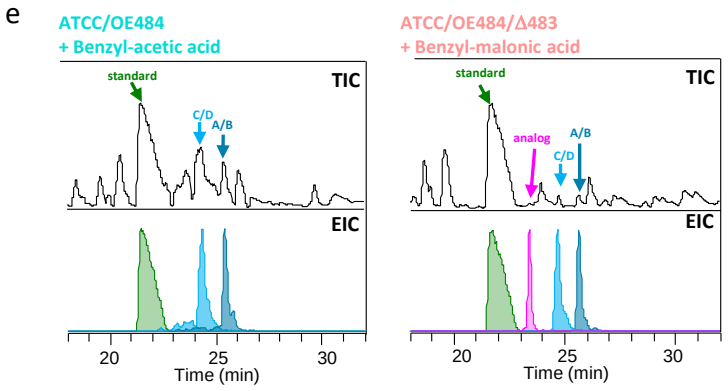




Approach	Compound	Amount relative to internal standard
PDB	Stambomycins	64%
	Analogue	n.d.
Mutasynthesis	Stambomycins	11.7%
	Analogue	7%



Approach	Compound	Amount relative to internal standard
PDB	Stambomycins	84.6%
	Analogue	11.4%
Mutasynthesis	Stambomycins	0.77%
	Analogue	138%



Approach	Compound	Amount relative to internal standard
PDB	Stambomycins	132%
	Analogue	n.d.
Mutasynthesis	Stambomycins	9.9%
	Analogue	0.08%

Figure 3.11 Quantitative analysis of PDB and mutasynthesis-mediated biosynthesis of stambomycin analogues. a) PDB: the culture medium is supplemented with monoacids, which are converted to extender units for AT_{12} by the two-enzyme pathway involving SamR0482 and SamR0483. **b**) Mutasynthesis: the culture medium is supplemented with diacids, which are converted to acyl-CoA extender units for AT_{12} by MatB_Sa. **c–e**) Quantitative analysis of the PDB and mutasynthesis crude extracts. The amount of stambomycins (C/D+A/B) and stambomycins analogues relative to the internal standard (Ery) were determined by the integrated area of EIC peaks (mean value of three individual fermentation results) and the peak aera of standard was taken as 100%.

3.3 Discussion

In contrast to many AT domains which exhibit strict substrate specificity, the AT_{12} domain of the stambomycin biosynthetic pathway naturally recognizes at least six atypical malonate extender units, testifying to its innate broad substrate tolerance. To enable a mutasynthesis approach aiming to exploit this intrinsic promiscuity, we aimed to disable production of these unusual starter units by targeting the enzymes (SamR0482 and SamR0483) involved in their biosynthesis from medium-chain fatty acids [40]. By deletion of the individual genes, we demonstrated that SamR0483 was critical for stambomycin biosynthesis in the absence of additional growth modification, but that SamR0482 could be complemented by other enzymes encoded in the genome. We therefore carried out mutasynthesis using the SamR0483 inactivation mutant.

Through the supplementation of the ATCC/OE484/Δ483 mutant with suitable precursors, 6 novel stambomycin analogues were obtained at varying yields. The fed compounds consisted of a series of malonic acids varying in the substitution pattern at the α-position, which were converted into their corresponding CoAs by a MatB-type enzyme (MatB_Sa), recognized by AT_{12} of the PKS, and then extended by the downstream modules. Analogues obtained in this way incorporated ethyl, butyl, allyl, benzyl, phenoxypropyl and 6-bromohexyl groups at C-26 instead of the native chains (**Figure 3.7**) (the results with feeding of octyl-malonic acid were ambiguous). For each of the analogues except that bearing the 6-bromohexyl group, a proportion was hydroxylated at C-28 by the P450 enzyme SamR0478, but non-hydroxylated compound was also obtained. Globally, these results reveal the tolerance of the stambomycin pathway (at least downstream of module 11) to unnatural building blocks/chain extension intermediates. Nonetheless, despite the

intrinsically high specificity of AT_{12} and the overall PKS system, not all of the alternative building blocks were accepted. Intriguingly, despite the hexyl/heptyl side chains of the native extender units, the non-native analogues showing the highest level of incorporation were significantly shorter (i.e. butyl and allyl), while isopropyl and phenyl were notably not accepted.

The efficient incorporation of butyl-malonic acid is also notable for a second reason, because it suggests that MatB_Sa has unusually high activity towards this substrate. Indeed, MatB_cinna shows only weak tolerance towards this substrate *in vitro* [144] (an engineered mutant of MatB_Rt exhibited *in vitro* activity to this substrate [153]), an observation supported by the fact that expression of MatB_cinna alongside MatB_Sa did not improve incorporation of butyl-malonate by the PKS.

In light of the broad substrate specificity of MatB_Sa and the AT_{12} domain, the two could in principle be leveraged together as a tool-kit for incorporation of alternative building blocks into other PKS assembly lines. Indeed, we have shown by comparative analysis that mutasynthesis is more efficient in *S. ambofaciens* than the alternative PDB approach, even though mutasynthesis did result in continued production of the wild type stambomycins.

In addition, it may be possible to further broaden the substrate tolerance of both enzymes via structure-guided mutagenesis. For example, an engineered version of MatB_Rt (T207G/M306I) showed improved catalytic activity toward ethyl-, propargyl- and allyl-malonic acids, as well some tolerance towards isopropyl-, butyl-, phenyl-, and azido-malonic acids [153]. The two residues corresponding to T207 and M306 in MatB_Sa are V190 and M293 (**Appendix Figure 3.2**), which could be modified in attempts to similarly broaden the substrate scope of MatB_Sa. The AT_{12} domain would also be an attractive target, as despite the presumed efficient provision of propargyl-malonic acid to the domain by MatB_cinna [153], none of the corresponding stambomycin analogue was obtained. Attempts to expand the substrate tolerance of AT_{12} could be guided by multiple available AT structures [106, 143] [170, 171]. Indeed, as reported in two recent works [106, 181], the structure of several AT-substrate complexes were determined by X-ray crystallography or molecular dynamic stimulation, which identified the specificity

determinants of AT domain in addition to the conserved motif (e.g. HAFH conserved for malonyl-CoA specific AT, YASH conserved for methylmalonyl-CoA specific AT) and hence revealed the critical information between the α substituents of the extender units and ATs. Site-directed mutagenesis of these specificity determinants has been proofed successfully to broaden or switch the substrate specificity of AT domains, although how this activity translates into the *in vivo* situation was not tested. However, it may also be necessary to develop higher-throughput methods for testing AT variants, as it is not always possible to identify on the basis of high-resolution structures alone, amino acids which can have important effects on enzyme activity [182].

Finally, it is worth mentioning that we are at the stage of purifying the allyl- and butyl-substituted stambomycins, in order to both conclusively prove their structures, and evaluate their biological activity relative to the parental structures. Given the promising anti-proliferative properties of the stambomycins, we have registered these novel structures in a declaration of invention with the SATT-Sayens, as a prerequisite to obtaining patent protection.

3.4 Materials and methods

3.4.1 Materials

All reagents and chemicals including the fed mono- and diacids were obtained from Sigma-Aldrich. Oligonucleotide primers were synthesized by Eurogentec.

The propargyl malonic acid (by Dr. Yaouba SOUAIBOU) and 6-bromohexyl malonic acid (by Yi YU Group, China) were chemically synthesized as described in reference [144].

Table 3.2 List of strains and mutants used in this chapter

Strains/mutants	Description and use (resistance)	Reference
Streptomyces		
ATCC/OE484	*S. ambofaciens* ATCC23877 carrying the integrative plasmid pIB0484 (strain overexpressing the LAL regulator SamR0484)	[40]
ATCC/OE484/Δ482	PCR targeting mutant of ATCC/OE484 with the gene *samR0482* replaced by a "*FRT+aadA+oriT+FRT*" cassette ($Spec^R$, $Apra^R$)	[180]
ATCC/OE484/Δ483	PCR-targeting mutant of ATCC/OE484 with the gene *samR0483* replaced by a scar ($Apra^R$)	This study

ATCC/pIB139/Δ483	PCR-targeting mutant of ATCC/pIB139 with the gene *samR0483* replaced by a scar ($Apra^R$)	This study
ATCC/OE484/Δ483::*samR0483*	ATCC/OE484/Δ483 complemented with the conjugative and integrative plasmid pOSV809-PermE_SamR483 (Kan^R, $Apra^R$)	This study
ATCC/pIB139/Δ483::*samR0483*	ATCC/pIB139/Δ483 complemented with the conjugative and integrative plasmid pOSV809_PermE_SamR483 (Kan^R, $Apra^R$)	This study
ATCC/OE483/Δ483/MatB_cinna	ATCC/OE484/Δ483 complemented with the conjugative and integrative plasmid pRT801_lacZ_PermE_MatB_cinna ($Spec^R$, $Apra^R$)	This study
E. coli		
DH5α	General cloning strain	[183]
BW25113/pKD20	PCR-targeting mutagenesis strain containing a λRED recombination plasmid pKD20 (Amp^R)	[184]
ET12567/pUZ8002 (ETU)	Non-methylating strain containing a mobilization plasmid for conjugation with *Streptomyces* (Kan^R, Cm^R)	[185]
S17-1	The conjugative donor strain containing the *tra* gene on the chromosome	[186]

Table 3.3 List of BACs and plasmids used in this chapter

BACs/plasmids	**Properties and use (resistance)**	**Reference**
BAC3	BAB19ZF4 from the genomic library of *S. ambofaciens*, in which the chloramphenicol resistance gene is replaced by a spectinomycin resistance gene ($Spec^R$)	This study
BAC3_K7Δ483	Mutant of BAC3 with gene *samR0483* replaced by a *"FRT+aadA+oriT+ FRT"* cassette ($Spec^R$, $Apra^R$)	This study
pIB139	Conjugative and integrative vector (φC31 *attP-int*, *ermEp**) ($Apra^R$)	[187]
pOE484	pIB139+*samR0484* ($Apra^R$)	[40]
pIJ778	Origin of "*FRT+aadA+oriT+FRT*" cassette, ($Apra^R$)	[188]
pIJ773	Origin of "*FRT+aac(3)IV+oriT+FRT*" cassette, ($Apra^R$)	[188]
pUWL-*oriT-flp*	Conjugative plasmid for "*FRT+aac(3)IV +oriT +FRT"* cassette excision (recombinase FLP) (Hyg^R)	[189]
pOSV809	Conjugative and integrative vector (φBT1 *attP-int*) (Kan^R)	[190]
pOSV809_PermE_SamR483	For complementation of the Δ*samR0483* mutant strain (Kan^R)	This study
pRT801_lacZ_PermE_MatB_cinna	Overexpression of MatB_cinna from *Streptomyces cinnamonensis* ($Spec^R$)	Gift from F Schulz, Ruhr-Universität Bochum, Germany

Table 3.4 List of primers used in this chapter

Oligo name	Sequence	Usage
Spec_For	GAGTTATCGAGATTTTCAGGAGCTAAG GAAGCTAAAATGAAGTCTACACGAAC CCTTTG	Replacement of the chloramphenicol gene by spectinomycin
Spec_Rev	AGTGAGCTAACTCACATTAATTGCGTT GCGCTCACTGCCCTTATTTGCCGACTA CCTTGG	Replacement of the chloramphenicol gene by spectinomycin
D483_For	GGAGGCAGGGTCGTCGTGTTGGAAGGT AGGGCTGGTATGATTCCGGGGATCCGT CGACC	Deletion of *samR0483*
D483_Rev	GCCCCCGTACAGGAGCGGCCGGCACG ACGAACGACGTCATGTAGGCTGGAGCT GCTTC	Deletion of *samR0483*
F483_For	CGGTTCCGTGCCGTCCGATA	Verification of deleted *samR0483*
F483_Rev	AGACCTCCAGCGGCAAGGTG	Verification of deleted *samR0483*
ermEp_For	TGCGGCCGCTGCTAGCCGAGTGTCCGT TCGAGTGGCGGCTTG	For amplification of *ermEp**+RBS
ermEp_Rev	GGTCCTCCTGTGGAGTGGTGTGGATCC TACCAACCGGCACGATTG	For amplification of *ermEp**+RBS
Comp483_For	CACCACTCCACAGGAGGACCATGTCGC TCCAGGAGCCTGTCTCGC	For amplification of *samR0483*
Comp483_Rev	GCGGCCGCTACTAGTTCACAGGGGAAT GTTCCCGTGTTTGCG	For amplification of *samR0483*
Ver483_For	CAAATGTAGCACCTGAAGTC	Verification of plasmid pOSV809-PermE_SamR483
Ver483_Rev	GTTCGGCCCCTTTTTTGGCC	Verification of plasmid pOSV809- PermE_SamR483
pOSV_For	GACTTCGCCCATCATGCGCTC	Verification of mutant ATCC/OE483/Δ483/*samR0483* and ATCC/pIB139/ Δ483:: *samR0483*
pOSV_Rev	GTGCTCAACGGGAATCCTGCTC	Verification of mutant ATCC/OE483/Δ483/*samR0483* and ATCC/pIB139/ Δ483:: *samR0483*
φBT1-attB_For	GACCTTGTGCTTGGTCGTCTTC	Verification of mutant ATCC/ OE483/Δ483/*samR0483* and ATCC/pIB139/Δ483::*samR0483*
φBT1-attB_Rev	GTAGATCGACAGGGCCATCCAC	Verification of mutant ATCC/ OE483/Δ483/*samR0483* and ATCC/pIB139/Δ483::*samR0483*

Key: "NNNNNNN" and "NNNNNNN" represent sequences identical to those at the right or left ends of the disruption cassette in the PCR-targeting system. "GCTAGC" is the restriction recognition site of *Nhe*I, and "ACTAGT" is that for *Spe*I.

3.4.2 PCR-targeting based genetic engineering to generate the mutant ATCC/OE484/Δ483

To render the BAC BAB19ZF4 proficient for selection following conjugation, its chloramphenicol resistance gene was replaced using a PCR-targeting approach, by a spectinomycin resistance gene cassette sourced from pIJ778 [188], resulting in BAC3 (**Table 3.3**). In order to obtain an in-frame deletion of the *samR0483* gene, the primer pair D483_For/D483_Rev (**Table 3.4**) was designed to PCR amplify the "*FRT+aac(3)IV+oriT+FRT*" disruption cassette from pIJ773 [188], affording PCR fragment *K7*Δ483. The PCR fragment was then electro-transformed into *E. coli* BW25113/pKD20/BAC3, giving rise to the mutant cosmid BAC3_K7Δ483. Subsequently, BAC3_K7Δ483 was introduced into *E. coli* ET12567/pUZ8002 and then *S. ambofaciens* ATCC23877 *via* intergeneric conjugation. The resulting exconjugants (ATCC/Δ483_ *aac(3)IV+oriT*) were selected for apramycin resistance and spectinomycin sensitivity, attesting to successful double crossover. The correct mutations were confirmed by PCR and sequencing. Thereafter, the cassette "*aac(3)IV+oriT*" was excised by the FLP recombinase encoded by plasmid pUWL-*oriT-flp* as described previously [189], leaving an 81 bp scar sequence (mutant ATCC/Δ483). Successful removal of the cassette was verified by PCR and DNA sequencing. Finally, the LAL regulator overexpression plasmid pOE484 and the parental vector pIB139 which was used as a control, were transferred into mutant ATCC/Δ483, giving rise to the mutants ATCC/OE484/Δ483 and ATCC/pIB139/Δ483, respectively (**Table 3.2** and **Figure 3.11**).

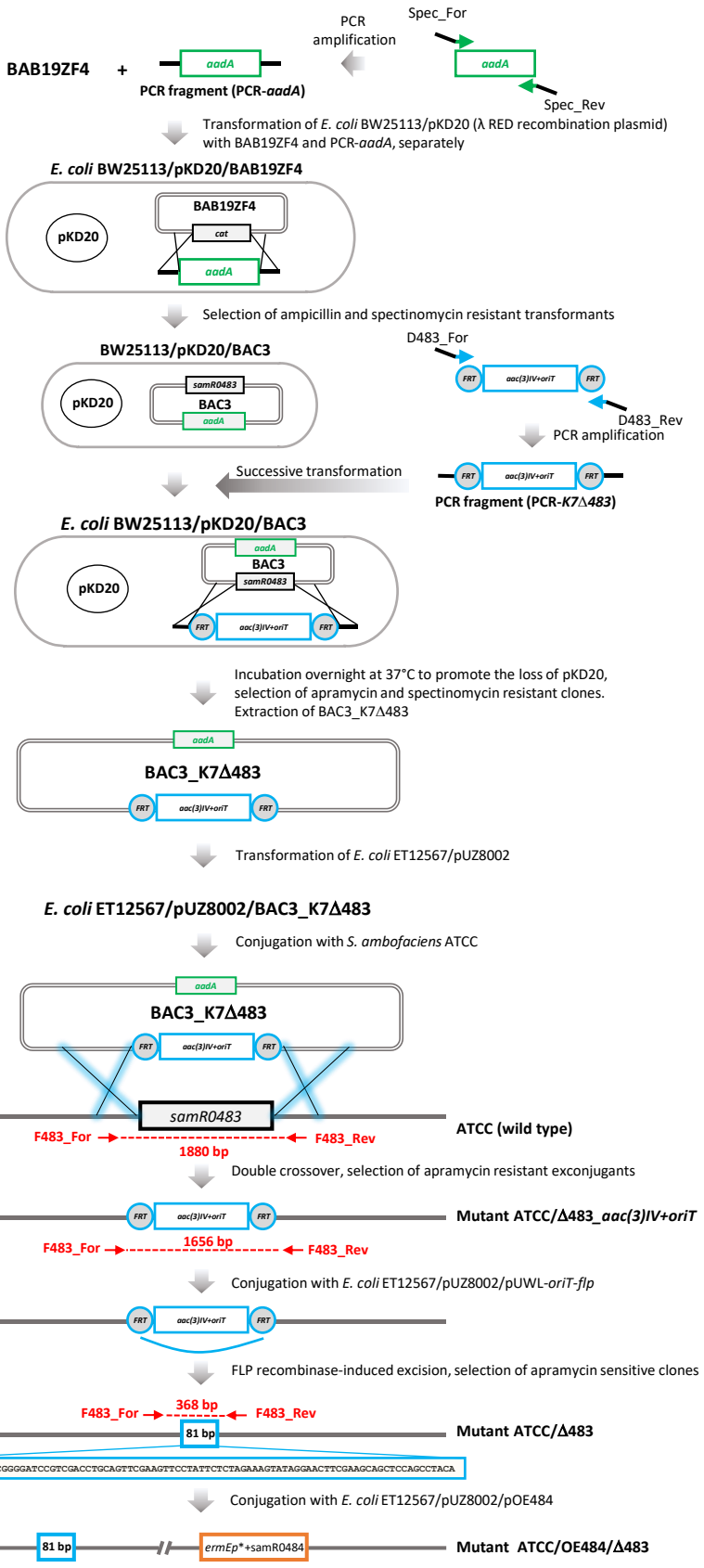

BAB19ZF4
+
aadA
PCR fragment (PCR-aadA)
PCR amplification
Spec_For
aadA
Spec_Rev
Transformation of E. coli BW25113/pKD20 (λ RED recombination plasmid) with BAB19ZF4 and PCR-aadA, separately
E. coli BW25113/pKD20/BAB19ZF4
pKD20
BAB19ZF4
cat
aadA
Selection of ampicillin and spectinomycin resistant transformants
BW25113/pKD20/BAC3
pKD20
samR0483
BAC3
aadA
D483_For
FRT
aac(3)IV+oriT
FRT
D483_Rev
PCR amplification
Successive transformation
PCR fragment (PCR-K7Δ483)
E. coli BW25113/pKD20/BAC3
aadA
BAC3
samR0483
pKD20
FRT
aac(3)IV+oriT
FRT
Incubation overnight at 37°C to promote the loss of pKD20, selection of apramycin and spectinomycin resistant clones. Extraction of BAC3_K7Δ483
aadA
BAC3_K7Δ483
FRT
aac(3)IV+oriT
FRT
Transformation of E. coli ET12567/pUZ8002
E. coli ET12567/pUZ8002/BAC3_K7Δ483
Conjugation with S. ambofaciens ATCC
aadA
BAC3_K7Δ483
FRT
aac(3)IV+oriT
FRT
samR0483
F483_For
F483_Rev
1880 bp
ATCC (wild type)
Double crossover, selection of apramycin resistant exconjugants
FRT
aac(3)IV+oriT
FRT
Mutant ATCC/Δ483_aac(3)IV+oriT
F483_For
1656 bp
F483_Rev
Conjugation with E. coli ET12567/pUZ8002/pUWL-oriT-flp
FRT
aac(3)IV+oriT
FRT
FLP recombinase-induced excision, selection of apramycin sensitive clones
F483_For
368 bp
F483_Rev
81 bp
Mutant ATCC/Δ483
ATTCCGGGGATCCGTCGACCTGCAGTTCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAAGCAGCTCCAGCCTACA
Conjugation with E. coli ET12567/pUZ8002/pOE484
81 bp
ermEp*+samR0484
Mutant ATCC/OE484/Δ483

Figure 3.12 PCR-targeting workflow for constructing the mutant ATCC/OE484/Δ483. The PCR-targeting manipulation was based on the "REDIRECT system" [188]. The employed cassette contained a resistance gene as well as *oriT*, flanked by FRT sites (FLP recognition targets), which allowed for FLP-mediated cassette excision. As described previously, the FLP recombinase carried on plasmid pUWL-*oriT-flp* was introduced into *S. ambofaciens via* intergeneric conjugation. After the excision, the marker-free mutant contained an 81 bp scar. The horizontal arrows indicate primer locations which were used to identify the wild type and double crossover progeny after cassette excision.

3.4.3 Complementation of mutant ATCC/OE484/Δ483

The gene *samR0483* and the constitutive promoter *ermEp** (including an RBS) were PCR amplified using primers ermEp_For/ermEp_Rev and Comp483_For/Comp483_Rev using ATCC genomic DNA and pIB139 as template, respectively. Two PCR fragments were fused into one fragment (*ermEp*+RBS+samR483*) using overlap extension PCR, prior to cloning into the vector pCR^{TM}-Blunt. The fragment was then excised from pCR^{TM}-Blunt by FD *Nhe*I and FD *Spe*I digestion, and then cloned into pre-digested (FD *Nhe*I and FD *Spe*I) and dephosphorylated (FastAP) vector pOSV809, giving rise to the construct pOSV809-*ermEp*-samR483.* The vector was then transformed into *E. coli* S17-1 and introduced into mutants ATCC/Δ483/OE484 and ATCC/Δ483/pIB139 via intergeneric conjugation. Kanamycin and apramycin resistant conjugants were selected and verified by PCR with primers pOSV-For/ φBT1-attB-Rev and pOSV-Rev/ φBT1-attB-For.

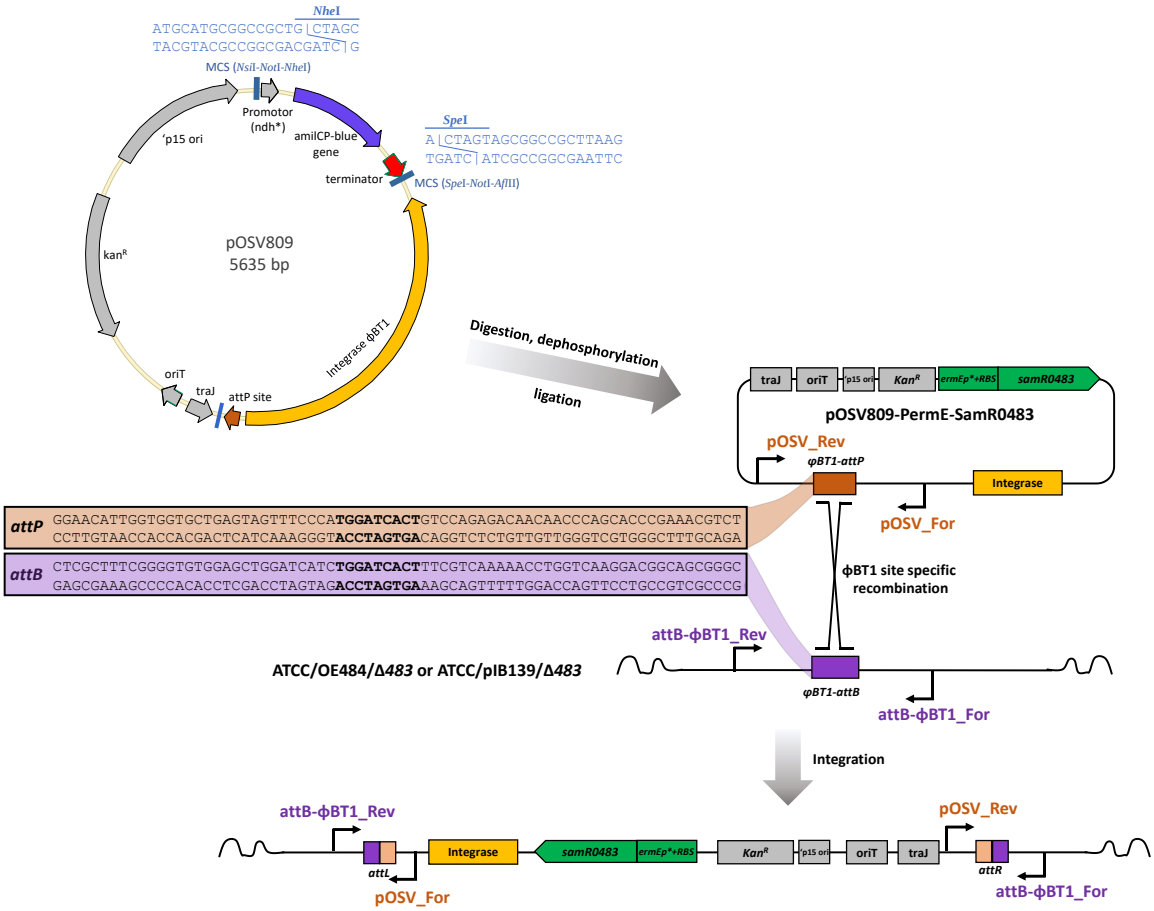


Figure 3.13 Schematic representation of complementation and the relevant verification strategy. The integration plasmid pOSV809 contains a gene encoding amilCP blue chromoprotein, which facilitates the selection of positive transformants upon successful incorporation of PCR fragment "*ermEp*+RBS+samR483*". The resulting plasmid pOSV809_PermE_SamR0483 was further confirmed by sequencing with the primer pair Ver483_For/Ver483_Rev. Thereafter, it was introduced into *Streptomyces* strains ATCC/OE484/Δ483 and ATCC/pIB139/Δ483 via conjugation with *E. coli* S17-1 (instead of ET12567/pUZ8002 due to a resistance incompatibility). The φBT1 *attP-int* locus on the plasmid enabled a site-specific integration with the counterpart *attB* in the *Streptomyces* strains. The sequences of the φBT1 *attP* and *attB* sites are shown in boxes, with bases that are identical between the sites (which represent the crossover region) highlighted in bold. Primer pairs attB-φBT1_For/Rev, attB-φBT1_For/pOSV_Rev and attB-φBT1_Rev/pOSV_For were used to verify correct exconjugants.

3.4.4 HPLC-MS analysis of fermentation metabolites

The methods described in **Chapter II** were employed for the processing of *Streptomyces* fermentation broths and HPLC-MS analyses. Quantification was achieved by adding erythromycin (1 mM) as internal standard during the extraction process. Specifically, 0.1

mg erythromycin was added for each 2 g of mycelia, with the resulting mixture resuspended in 140 μL of methanol (70 μL methanol/1 g of cell pellet). Alternatively, the yield of stambomycin analogues were determined using the stambomycin standard curve (**Chapter II**), based on the integrated areas of EIC (extracted ion chromatogram) peaks.

Chapter IV Identification of desferrioxamine derivatives from *S. ambofaciens*

This chapter reports a new strand of research originating from the two previous projects. In the course of metabolite analysis of the DDs mutants and mutasynthesis mutant, a series of metabolites unrelated to stambomycins were identified, which are surprisingly related to desferrioxamine siderophores (DFOs) and several novel forms of DFOs based on the molecular network generated by the Global Natural Products Social Molecular Networking-platform (GNPS).

4.1 Introduction

Iron is an essential nutrient element for most organisms. It participates in many biological metabolic processes, including photosynthesis, respiration, nitrogen fixation and DNA biosynthesis [191]. Although iron is the most abundant transition metal element in the earth's crust, its bioavailability is generally low. This poor availability derives from the fact that iron exists mainly in its oxidized ferric (Fe^{3+}) state, e.g., in the form of trivalent oxide or hydroxide, which is extremely insoluble (solubility product of ~10^{-39} at physiological pH) under the natural aerobic conditions suitable for most bacteria. To tackle the iron shortage problem, bacteria have developed a variety of mechanisms to acquire this essential element from the environment, notably including the synthesis and utilization of siderophores. Siderophores are a class of low-molecular-weight specialized metabolites capable of scavenging iron from environmental stocks by forming soluble Fe^{3+} complexes, which are then actively taken up via specific receptors.

There is an extensive body of work concerning siderophores, and in particular, their chemical structures and the molecular mechanisms underlying their synthesis, export, uptake and regulation [191-194]. Siderophore often functions as virulence factors in many bacterial pathogens (e.g. aerobactin in enteric bacteria [195]) due to absorption of iron from the host necessary for pathogen reproduction and infection. On the contrary, siderophores produced and secreted by microbes can also play an important role in the control of pathogens, because iron chelation can inhibit the use of iron by other organisms, thereby inhibiting the growth and metabolic activity of pathogenic bacteria (e.g. *Lysinibacillus.*

sphaericus ZA9 exhibts high antagonistic activity against several pathogens due to its high production of siderophores [196]). Furthermore, siderophores can act as chemical cues in interspecies interactions. For example, siderophores from a variety of microbial species were shown to alter cellular development and accelerate sporulation in *Bacillus subtilis* [197]. In addition, siderophores produced by one species can be taken up by another species ('siderophore piracy') or can induce the production of endogenous siderophores under iron-limited conditions as shown with marine bacteria [198, 199]. Alternatively, siderophores can be used for the common good. This is the case, for example, with pyoverdine, a fluorescent siderophore produced by certain pseudomonads which can be secreted from the producer and exploited by non-producing pseudomonads [200]. Together, these various modes of siderophore action help to shape competition for iron and facilitate the different types of social interactions in soil bacterial communities [201].

Siderophores are selective for iron, but can also bind a range of metals including zinc, manganese, copper, nickel, gallium, and aluminum [202]. Cases have also been reported in which siderophores function as multipurpose metallophores (e.g. yersiniabactin scavenges zinc, copper and nickel in addition to iron) [203, 204]. On the basis of the moiety involved in iron chelation, siderophores are classified into four basic types: catecholate, hydroxamate, carboxylate and the mixed type siderophores [205] (**Figure 4.1**). Desferrioxamines (DFOs) are well-known members of the hydroxamate siderophore family, which are synthesized and excreted by many *Streptomyces* species. Among these, desferrioxamine B is of clinical importance, as it is approved by the Food and Drug Administration (FDA) to treat metal poisoning. These diverse structures are assembled via two main pathways: non-ribosomal peptide synthetases (NRPS) [206] and NRPS-independent siderophore (NIS) pathways [207]. Notably, the siderophore products of NIS pathways often contain diamine, citric acid, and dicarboxylic acid building blocks.

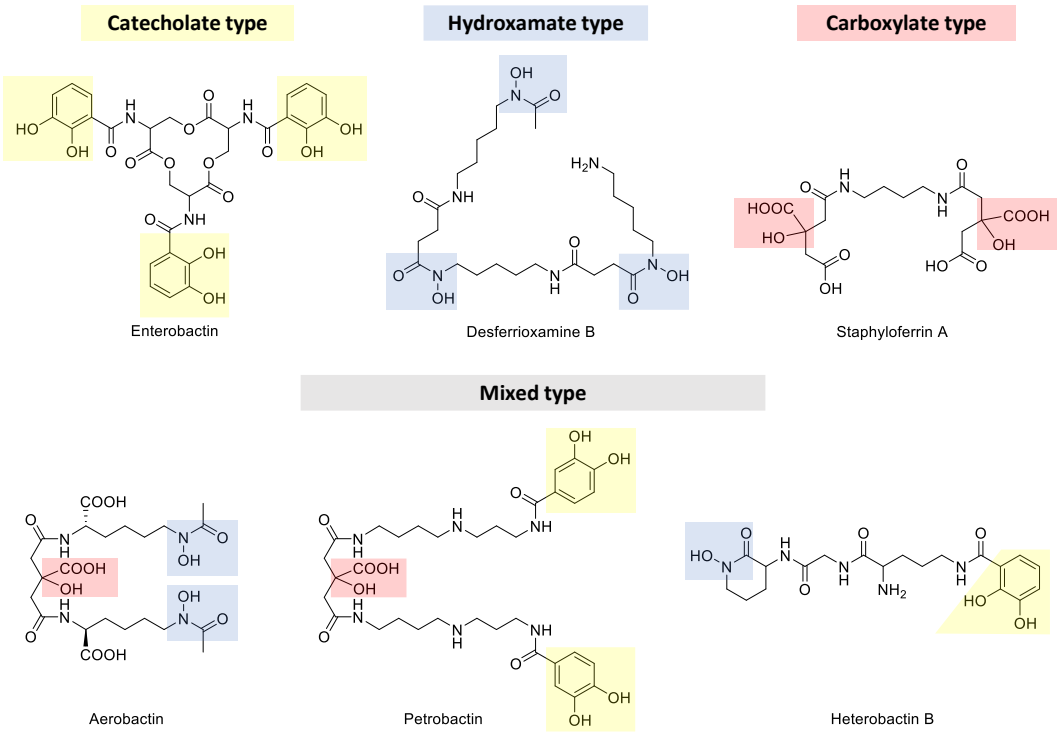


Figure 4.1 Representative examples of the four chemical classes of bacterial siderophores. The four types are classified based on the functional groups involved in iron chelation. The catecholate type, as exemplified by enterobactin produced by *Escherichia coli*, is a siderophore in which the hydroxyl group of the catechol rings binds ferric iron [208]. The hydroxamate type, represented here by desferrioxamine B produced by *Streptomyces coelicolor*, is equipped with hydroxyl side groups that ligate ferric iron [209]. Carboxylate siderophores (also called citrate siderophores), as exemplified by staphyloferrin A produced by, contain hydroxy acid groups for iron chelation [210]. The mixed type siderophores, as illustrated here by aerobactin from *Aerobacter aerogenes* [195], petrobactin from *Marinobacter hydrocarbonoclasticus* [211], and heterobactin B produced by *Phodococcus erythopolis* [212], contain some combination of the previously-mentioned functional groups involved in iron chelation. The same siderophores can be produced by different species, while a single species can produce different types of siderophores.

The BGCs encoding siderophores are wildly distributed in *Streptomyces*, allowing them to respond to resource competition within the microbial soil community. *Streptomyces coelicolor* A3(2) has been well-studied in this context, and shown to produce coelichelin, DFO-B and E under iron-deficient conditions [209]. Coelichelin, a tetrapeptide siderophore, is encoded by a NRPS gene cluster (*cch*) located on the *S. coelicolor* chromosome [213]. In contrast, a non-NRPS cluster of six genes (*desABCDEF*) directs the biosynthesis of DFO-B and E (**Figure 4.2a**). DFO-B and E are biosynthesized from L-lysine, acetyl-CoA and succinyl-CoA as building blocks (**Figure 4.2a**). The first step

involves the generation of 1,5-diaminopentane (otherwise known as cadaverine) by decarboxylation of L-lysine, which is catalyzed by DesA, a pyridoxal 5-phophate-dependent L-lysine decarboxylase. This step is followed by mono-*N*-hydroxylation carried out by DesB, a FAD-dependent amine monooxygenase, resulting in *N*-hydroxy-cadaverine. This compound is then *N*-acetylated by DesC, an acyl CoA-dependent acyl transferase, to yield *N*-hydroxy-*N*-acetylcadaverine (HAC). Alternatively, DesC catalyzes *N*-succinylation to give *N*-hydroxy-*N*-succinylcadaverine (HSC). Three units of HSC are activated and condensed by DesD, an NRPS-independent peptide synthetase (NIS enzyme), to produce HSC-HSC-HSC, followed by ring-closing reaction, yielding DFO-B. DFO-E is composed of two units of HSC and a unit of HAC. The two other genes in the *des* cluster, d*esE* and *desF*, encode respectively a cell surface-associated lipoprotein receptor component of an ABC transporter involved in ferrioxamine uptake, and a putative ferrioxamine reductase, an enzyme that removes iron from hydroxamate siderophores.

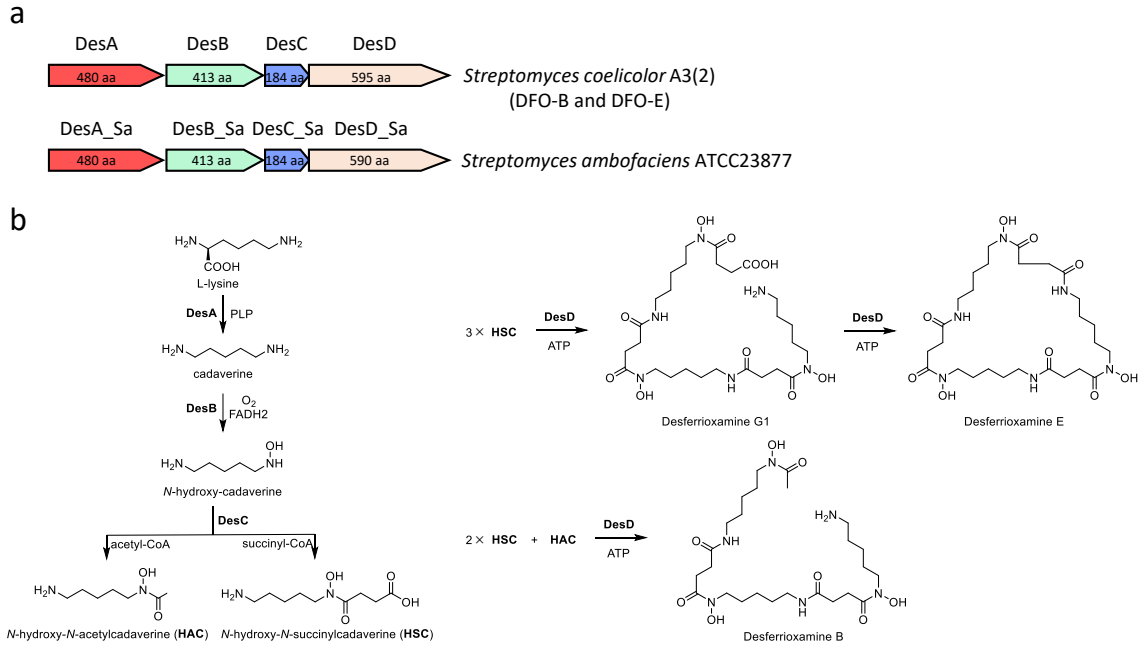


Figure 4.2 Representative biosynthetic gene clusters and pathway to NIS desferrioxamine. a) The BGC for desferrioxamines (E, G1, and B) in *S. coelicolor* [209], and the corresponding DFO BGC in *S. ambofaciens* (genes encoding transporter and regulatory enzymes are not shown). The same color indicates a shared enzymatic function, and the size of each protein in amino acids is indicated. **b**) Currently proposed biosynthetic pathway to DFO-B and E, based on ref [209].

Previous studies indicated that *S. ambofaciens* harbours a *des* BGC and a *cch* BGC, both of which show strong conservation with the corresponding clusters in *S. coelicolor*. [214, 215]. The six enzymes encoded by *desABCDEF* show respectively 95%, 91%, 83%, 91%, 92%, 87% amino acid sequence identity with their homologs in *S. coelicolor* ((**Figure 4.2a**, **Appendix Table 4.1**), and are responsible for the biosynthesis of DFO-B and E [215].

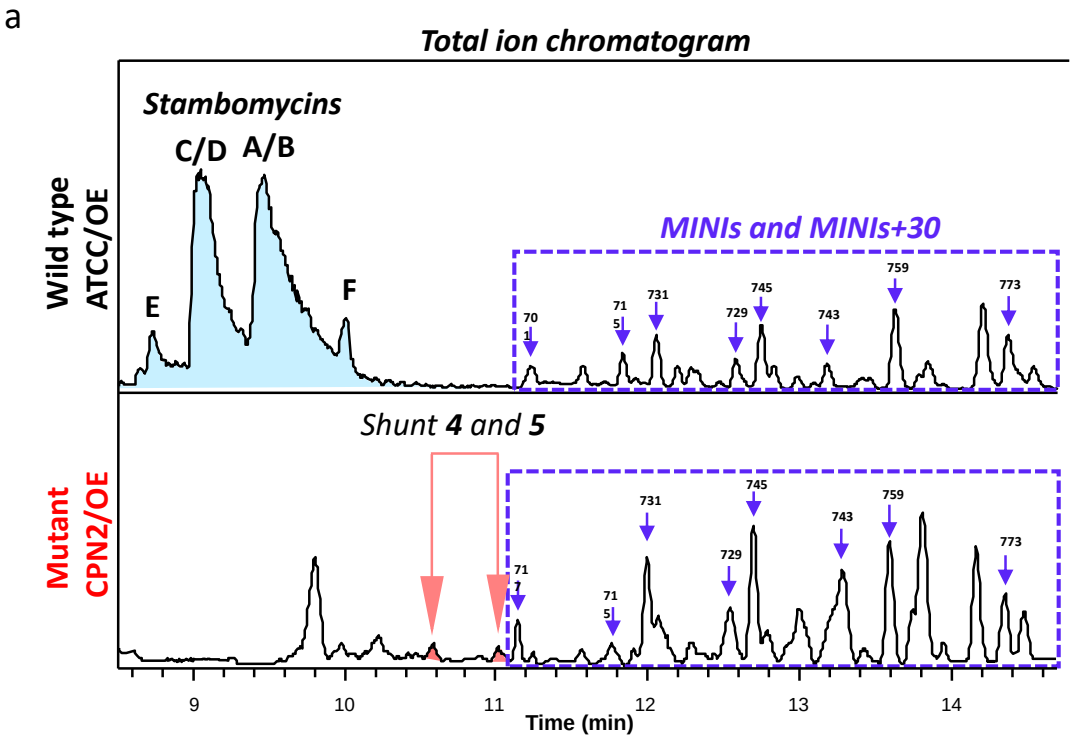
In this chapter, we report four series of novel siderophores, in addition to the previously identified DFO-B and E, produced by *Streptomyces ambofaciens* ATCC23877 and mutants obtained by genetic engineering of the stambomycin gene cluster. Among the four series of new DFOs, we can assign one as the acyl-DFOs, a class of recently identified siderophores produced by *S. coelicolor* when interacting with other actinomycetes [216]. Notably, the other three series of DFOs, termed acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30 were observed exclusively from *S. ambofaciens*. Their yields are comparable or even superior to acyl-DFOs, especially for acyl-DFOs+30. Furthermore, the latter two classes are proposed to be tetra- and penta-hydroxamate metal chelators distinct from the common bis-hydroxamate (e.g. bisucaberin [200]) or tris-hydroxamate iron (e.g. DFOs) chelators, implying the capability to bind high-valence metal ions and the enzymatic potential of NIS pathway for extended oligomerization.

4.2 Results

4.2.1 Identification of novel forms of desferrioxamines (DFOs) from *Streptomyces ambofaciens*

In the course of HPLC-MS analysis of DD mutants (**see Chapter II**), two novel series of related metabolites were identified with the masses of 687.5013, 701.5171, 715.5303, 729.5466, 743.5624 (referred to as "MINIs") and 717.4736, 731.4901, 745.5048, 759.5206, 773.5753 (referred to as "MINIs+30") (highlighted in blue dashed box in **Figure 4.3a**) The two series of masses are separated from each other by 30 mass units and differ sequentially from each other within a series by +14. Initially, we hypothesized that the MINIs could correspond to the polyketide products (shunt **4** and **5**) released form Pks4, followed by loss of a water molecule, either by dehydration or cyclization (the predicted cyclic form is shown in **Figure 4.3b**). However, these compounds were also identified in the wild type

strain, albeit at lower levels (**Figure 4.4**), as well as in a *S. ambofaciens* mutant strain lacking the first stambomycin PKS subunit (ATCC/OE484/Δ467) (**Appendix Figure 4.1**) and in the mutant ATCC/OE484/Δ483 used for the mutasynthesis approach (Chapter XX, **Figure 4.4**). Therefore, these data together argue that the MINIs and MINIs+30 belong to a group of new compounds with no relationship to the stambomycins.



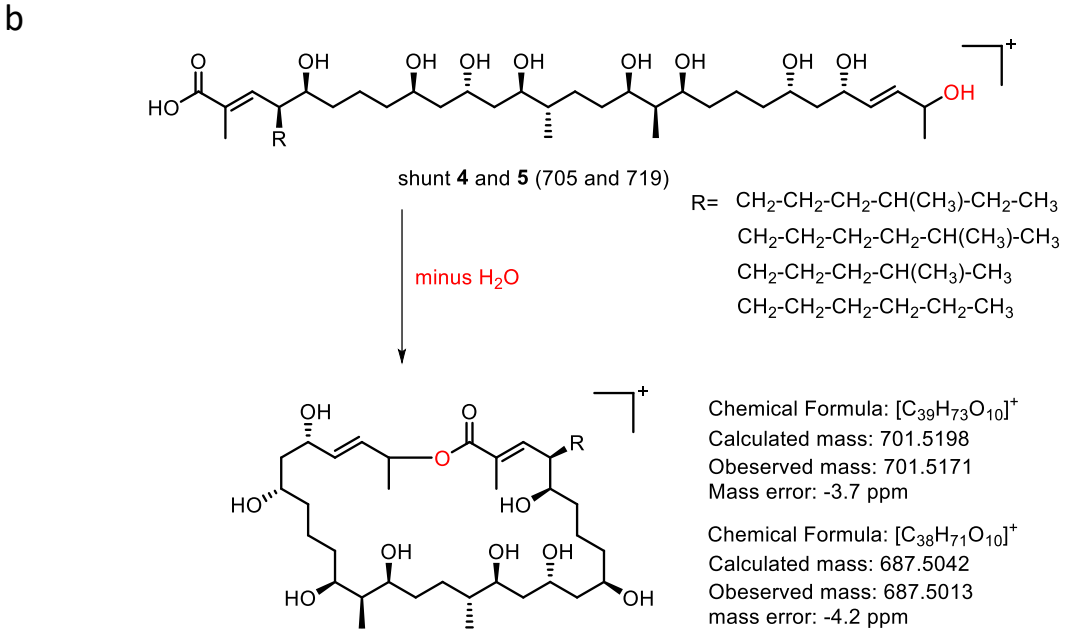


Figure 4.3 Two series of new metabolites detected from *S. ambofaciens* wild type and DD engineering mutant strains. a) Total ion chromatogram of the whole extract from mycelia of wild type strain and DD mutants (represented by mutant CPN2/OE484). **b**) The predicted structure of 687 and 701, corresponding to the cyclized form of the shunt products **4** and **5** (mass are 705 and 719) from Pks4. The side chains of 701 and 687 presented here are identical to that of stambomycins A–D, derived from the unusual extender units recruited by AT_{12}. We initially hypothesized that the remaining MINIs (715, 729 and 743) could be the cyclized polyketides harboring R groups of different length. As for MINIs+30, we speculated that the proton of one of the "-CH_2-" groups in the MINIs was replaced by a methoxy group (i.e. -CH_2- → -$CH(OCH_3)$-). (Note: the discrepancy in retention time for all metabolites in this figure and **Figure 4.4** reflects the fact that analyses were carried out on two different machines, as described in **section 4.4.1.**)

In order of increasing retention time, the detected $[M+H]^+$ masses of the peaks in red (MINIs) in **Figure 4.4** are 673.4855, 687.5013, 701.5171, 715.5327, 729.5485, 743.5641, 757.5797. Similarly, the observed masses for peaks in green (MINIs+30) are 703.4599, 717.4756, 731.4912, 745.5067, 759.5225, 773.5383, 787.5539. Apart from MINIs and MINIs+30, we unexpectedly found two more clusters of related peaks (as represented by analysis of mutant K7N1/OE484): 903.5765, 917.5923,931.6079, 945.6235, 959.6350, 973.6550, 987.6705 (in blue) and 1103.6931, 1117.7086, 1131.7421, 1145.7396, 1159.7555, 1173.7711, 1187.7870 (in yellow), which differ sequentially from MINIs+30 by respectively +200 and +200+200, and as such are annotated as "MINIs+200+30" and "MINIs+200+200+30". There are 7 members for each mass series, and all members differ sequentially by "+14" within each series. In addition, the retention time for each MINI/MINI+30/MINI+200+30/MINI+200+200+30 group is quite close. These observations prompted us to hypothesize that they are all members of the same family of metabolites. However, on the basis of the integrated EIC peaks, we observed that the yields of the various series of compounds differed both within and between the various strains. Specifically, the yields of MINIs and MINIs+30 are always higher than MINIs+200+30 and MINIs+200+200+30. Furthermore, the yields of these metabolites from the DD engineering mutants are quite similar, but nonetheless 3.4- and 3.0-fold higher than from wild type ATCC/OE484 and strain ATCC/OE484/Δ483 used for mutasynthesis, respectively.

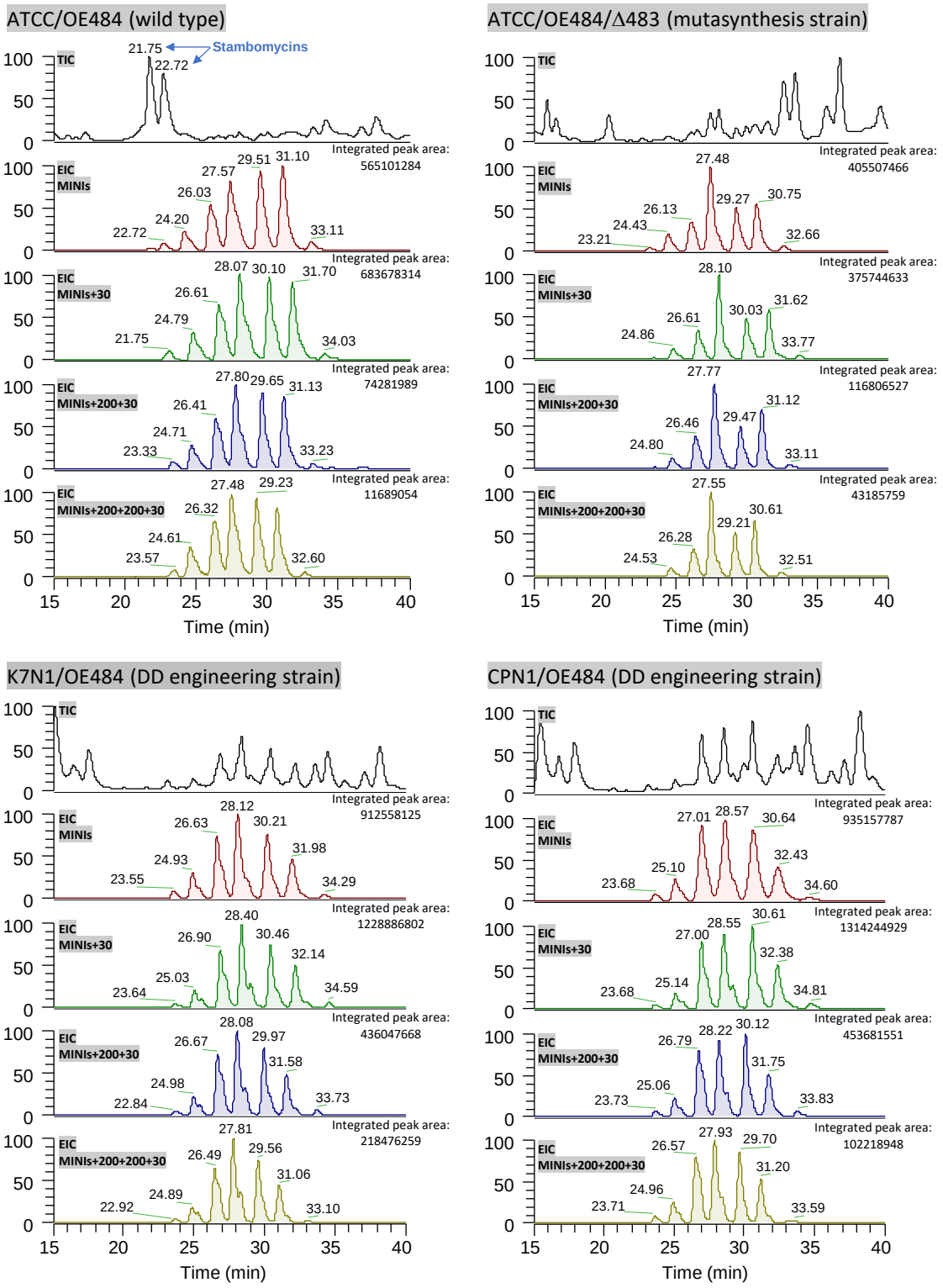


Figure 4.4 HPLC-MS analysis of 'MINI' and 'MINI' derivatives from wide type ATCC/OE484, mutasynthesis strain ATCC/OE484/Δ483 and DD engineering mutants. The DD engineering mutants are represented by K7N1/OE484 (PCR-targeting generated mutant) and CPN1/OE484 (the CRISPR-Cas9

generated mutant). TIC (total ion chromatogram) of the mutant extracts and the EIC (extracted ion chromatogram) of compounds corresponding to the MINIs, MINIs+30, MINIs+200+30, and MINIs+200+200+30 are shown. In each strain, the overall yields of MINI and MINI derivatives were estimated from the combined integrated EIC peak areas. One set of the representative chromatograms from two or three replicate experiments is shown, but the integrated area represents the average of replicate measurements, as detailed in **Appendix Table 4.2**.

Using the Global Natural Products Social Molecular Networking-platform (GNPS) [217], a molecular network of mutant K7N1/OE484 based on the fragmentation spectra was generated. Compounds, represented by nodes in the network, are sorted based on the similarity of their fragmentation spectra. Nodes representing highly similar spectra are connected by edges and are likely to be structurally similar. Notably, the molecular network revealed that the MINI series of compounds (shown with blue dots in **Figure 4.5**) corresponds to various acyl-desferrioxamine siderophores (acyl-DFOs). According to the related compound ID hits, we identified these blue dots as C9–C15 acyl-DFOs (**Figure 4.6**). In addition, their corresponding iron-cheated form, called ferrioxamines (acyl-FOs, the yellow dots in **Figure 4.5**), were also observed and linked to DFOs in the same subnetwork. Furthermore, bisucaberin (an *N*-hydroxy-*N*-succinyl diamine (HSD)-based siderophore like DFO, but which forms a cyclic dimer, bottom green dot), DFO-E (middle left green dot), DFO-D2 (middle right green dot), Desf-07 (turquoise dot [218]) and aluminum (Al)-chelated DFO-B (the green dot near yellow dots) were also observed and annotated in this subnetwork (**Figure 4.5**). Structures of all known compounds in this subnetwork are also shown in **Figure 4.6**.

Certain members of "MINIs+30" "MINIs+200+30" "MINIs+200+200+30" were also present in the same subnetwork and highlighted respectively in orange, pink and red with black outlines (**Figure 4.5** and **Figure 4.6**), proving that they are structurally related and belonging to the same metabolite family. These nodes are connected to DFOs and acyl-DFOs but without compound ID hits from the GNPS database, an observation which indicated that they are three novel series of DFO derivatives. We therefore renamed them as acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30, respectively (**Figure 4.5** and **Figure 4.6**).

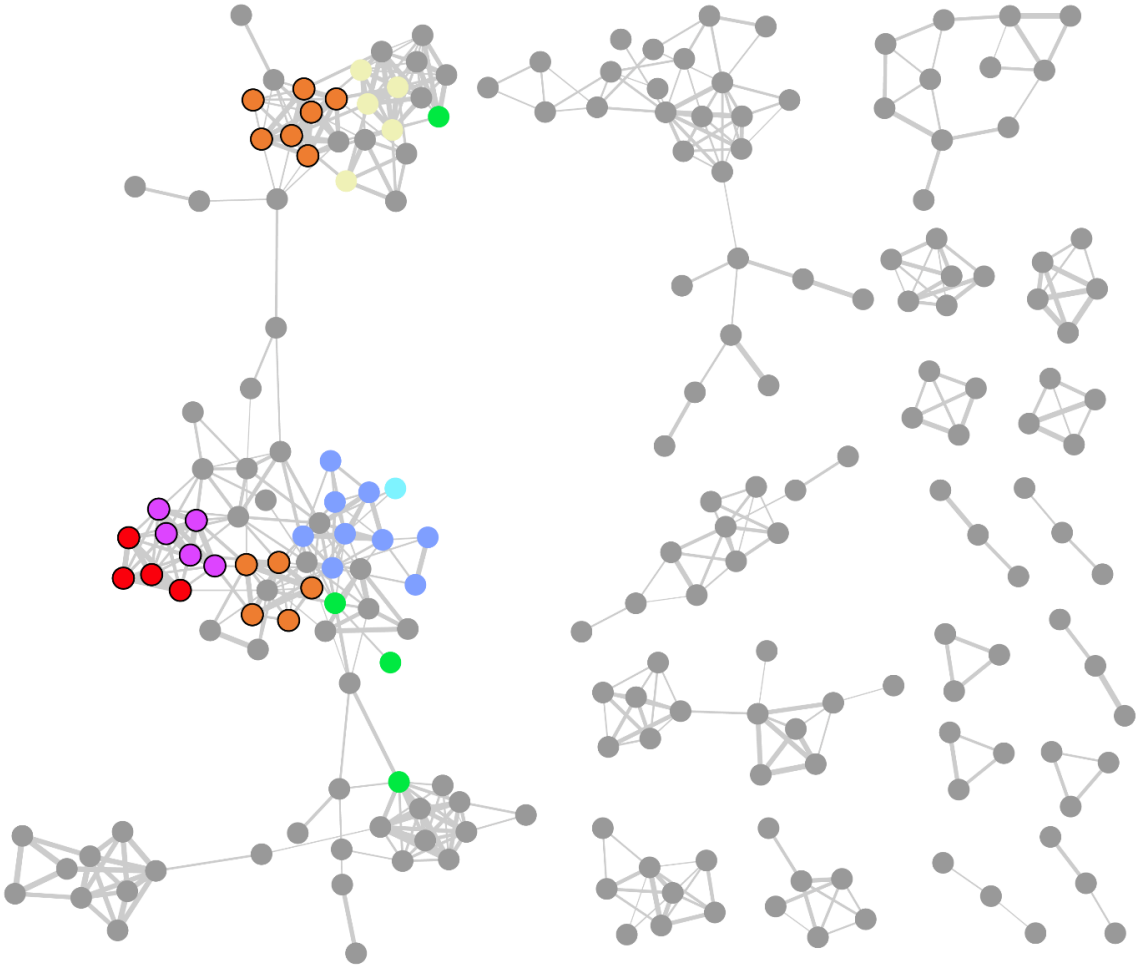


Figure 4.5 Molecular network of metabolites produced by K7N1/OE484. Subnetworks with nodes less than 3 members are not shown. Molecules are represented in nodes based on their protonated masses in Da and colored by group variants. Nodes are connected if the cosine similarity of fragment spectra is ≥0.7 (edge thickness reflects similarity score). The distance of the edge between two nodes corresponds to the relatedness of the two consensus spectra. The first subnetwork (left of the figure) is of interest as it is related to DFOs. Nodes shown by green dots are the well-known siderophores (DFO-E, D2, bisucaberin, and the Al-cheated forms of DFO-B). Nodes in blue dots indicate acyl-DFOs, varying in the length of the acyl-chain. Additional dots in orange, pink and red with black outlines correspond to acyl-DFOs+30 (and the corresponding iron-chelated form (also in orange) located near the yellow dots), acyl-DFOs+200+30 and acyl-DFOs+200+200+30, respectively.

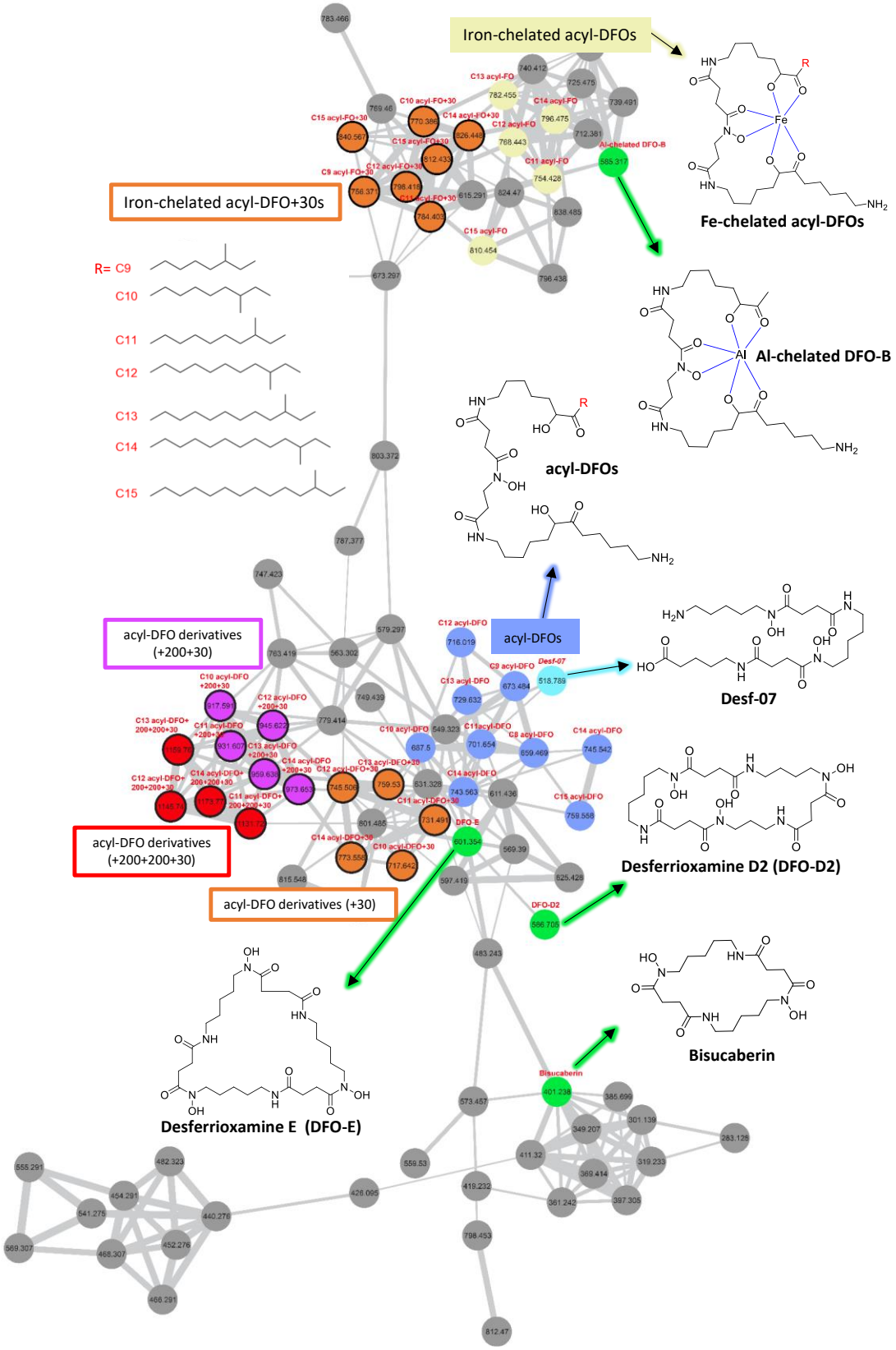

Iron-chelated acyl-DFOs
Iron-chelated acyl-DFO+30s
R= C9
C10
C11
C12
C13
C14
C15
Fe-chelated acyl-DFOs
Al-chelated DFO-B
acyl-DFOs
acyl-DFO derivatives (+200+30)
acyl-DFO derivatives (+200+200+30)
acyl-DFO derivatives (+30)
Desf-07
Desferrioxamine D2 (DFO-D2)
Bisucaberin
Desferrioxamine E (DFO-E)

Figure 4.6 Annotation of the subnetwork related to DFOs. The protonated mass in Da of each dot is shown in the center in this subnetwork. The names of all known compounds are given on the top of each dot, and their structures shown. The remaining members of acyl-DFOs+30 (C9 and C15), acyl-DFOs+200+30 (C9 and C15) and acyl-DFOs+200+200+30 (C9, C10 and C15) have not been presented in the network, due to their relatively low ion densities and lack of available MS/MS fragments, consistent with their low abundance, as shown in **Figure 4.4**.

4.2.2 Structure prediction of novel acyl-DFO derivatives

According to the previously elucidated structures and MS/MS fragment annotation of acylated DFOs [216], we found the MS/MS fragments 401.3, 319.3 and 201.1 are common to all 7 members (C9–C15) of acyl-DFOs (**Appendix Figure 4.2**, underlined by dashed lines in **Figure 4.7**). In addition, the other fragments of C12 acyl-DFO including 397.3, 515.6 and 597.6 correspond respectively to the fragments 411.3, 529.5, and 611.4 from C13 acyl-DFO with a difference of "+14" (an analogous "+14" difference was observed for all other members of the acyl-DFOs), indicating that these groups of fragments differ sequentially by a "$-CH_2$" unit among members of acyl-DFOs, and thus correlate to the acyl chain length (**Appendix Figure 4.2,** underlined by solid lines in **Figure 4.7**).

Interestingly, the three acyl-correlated fragments (397.3, 515.6 and 597.6) derived from the C12-acyl-DFO were also observed in the MS/MS pattern of the compounds with *m*/*z* of 745.5067, 945.6235 and 1145.7396. This observation suggested that the three compounds have the same C12 acyl-group, prompting us to name them as C12 acyl-DFO+30, C12 acyl-DFO+200+30 and C12 acyl-DFO+200+200+30. To facilitate further analysis, we analyzed all the detected MS/MS fragments for each compound, as well as annotated in detail the MS/MS spectra for each of the compounds sharing the same acyl group based on the MS/MS pattern of the C12 acyl-DFO (**Figure 4.7**). As shown in **Figure 4.7**, the set of common fragments (401.3, 319.3 and 201.1) defined above for C12 acyl-DFO, was detected in C12 acyl-DFO+30 as fragments with *m*/*z* 431.3, 349.3 and 231.2, thus differing from the first set of fragments by a common "+30" (the same observation could be made for the other members of acyl-DFOs+30, **Appendix Figure 4.3**). Taken together, these data strongly suggested that the "+30" difference occurred on the primary amine group ($-NH_2$) of the compounds.

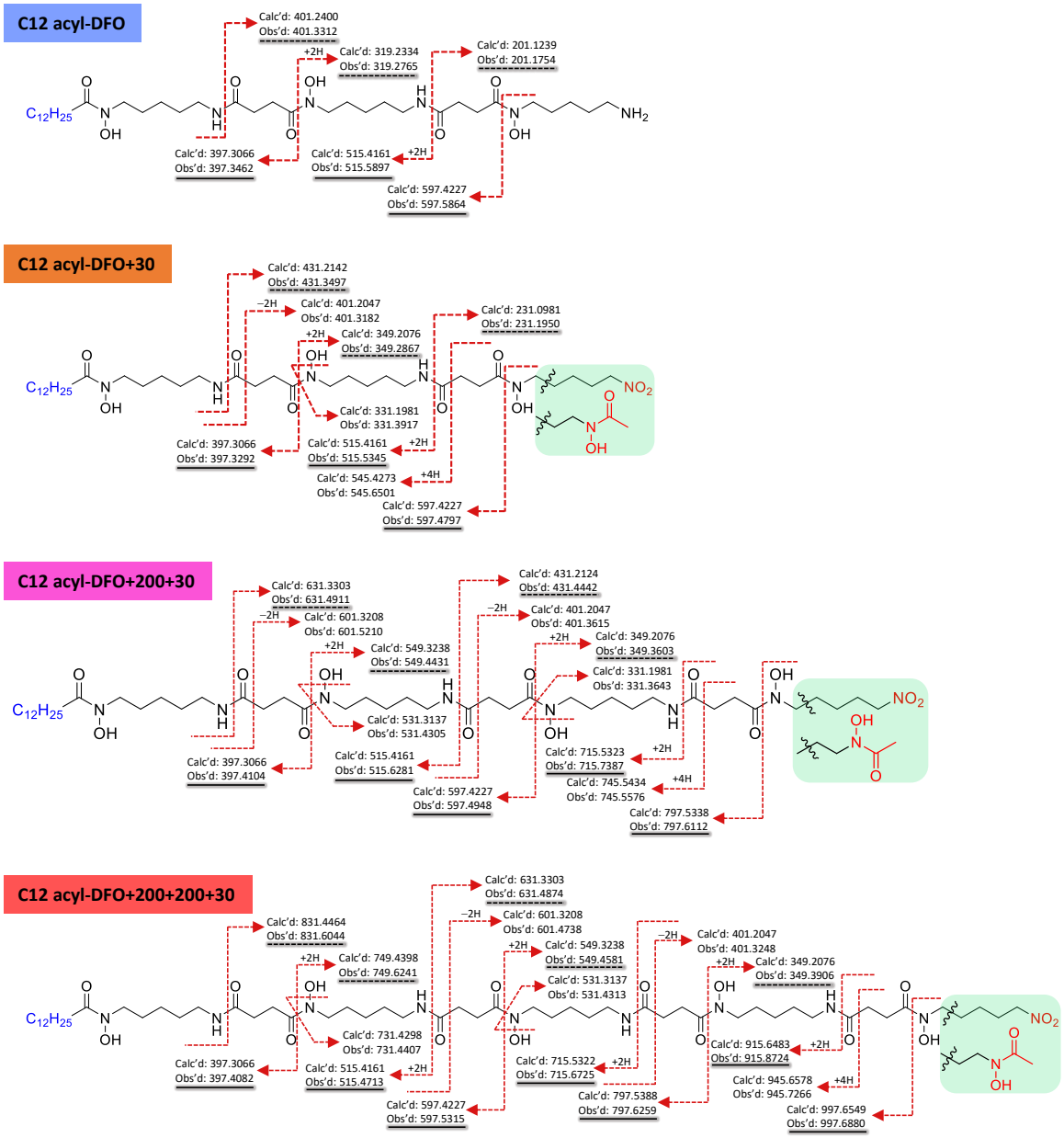


Figure 4.7 Summary and comparison of the observed HPLC-MS/MS fragmentation patterns of C12 acyl-DFO and its corresponding derivatives. The structure and MS/MS fragment assignment of C12 acyl-DFOs are based on previously-published work [216]. Accordingly, the structures of C12 acyl-DFO+30, C12 acyl-DFO+200+30 and C12 acyl-DFO+200+200+30 are predicted based on their common fragments (dashed underlines) and C12-correlated fragments (solid underlines) that they share with C12 acyl-DFO. The fragmentation leading to 401.3128, 331.3917 and 545.6501 in C12 acyl-DFO+30 was inferred based on the reported fragmentation patterns of Desf-06 [218] and acyl-DFO 846 [216].

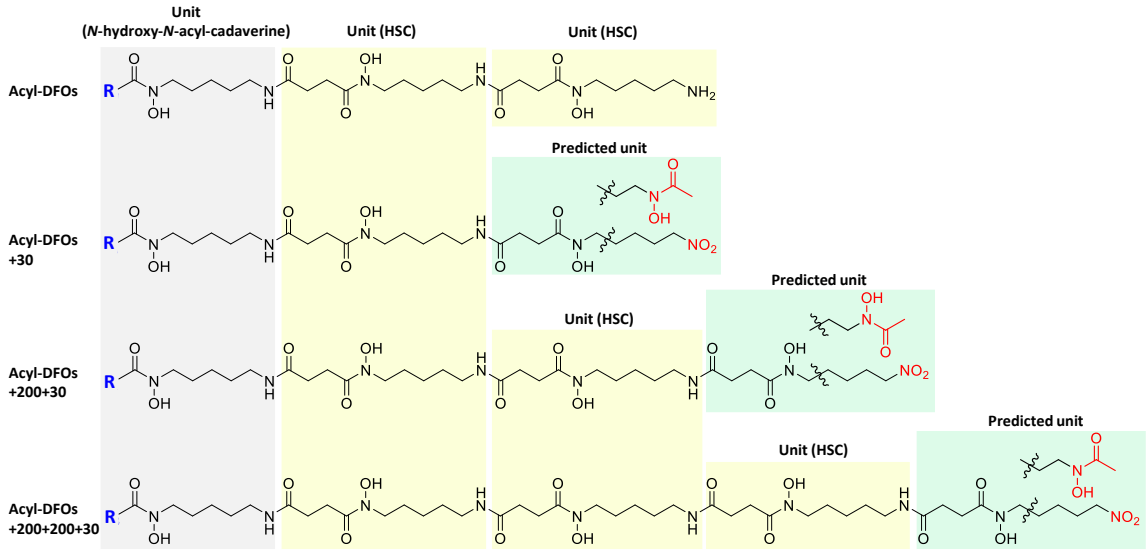


Figure 4.8 Comparison of the monomer composition of acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30. The structure of the acyl-DFOs is based on ref [216], while the structures of the additional derivatives were predicted on the basis of the MS/MS fragmentation patterns. Various units shown in different colors are linked together by amide bonds to form the various DFOs. The "+200" refers to incorporation of an additional HSC unit, while the "+30" is predicted to originate from one of two alternative modifications to the terminal amine group.

Although the genuine structures of the novel acyl-DFOs remains to be determined by NMR characterization, we can propose two alternative hypothesis to explain the origin of the "+30": i) transformation of the terminal amine group ($-NH_2$) to a nitro group ($-NO_2$), occurring during the LC-ESI-MS analysis (an idea proposed by my colleague Dr. Yaouba SOUAIBOU). This proposition is based on the reported observation that during ESI, electrochemical reactions can occur at the liquid-metal interface on the capillary, leading to reactive oxygen species (e.g. •OOH and H_2O_2) capable of oxidizing the amine [219, 220]. Alternatively, the NO_2 could also possible derived from enzymatic oxidation of NH_2 by SAM23877_8801, an [2Fe-2S] Rieske non-heme oxygenase showcasing 39% amino acid sequence identity to PrnD (**Appendix Figure 4.4**). PrnD is N-oxygenase involved in genuine nitro group formation from amine group, involved in the biosynthesis of the antifungal antibiotic pyrrolnitrin in *Pseudomonas fluorescens* [221, 222]; and, ii) use of an alternative 1,3-diaminopropane-derived building block during compound assembly (**Figure 4. 8**).

By comparison to the masses of acyl-DFOs+30, the "+200" present in the acyl-DFOs+200+30 and acyl-DFOs+200+200+30 was speculated to arise from incorporation of one or two additional building blocks of HSC, respectively (observed by my colleague Dr. Yaouba SOUAIBOU) (**Figure 4.8**). This hypothesis was further supported by analysis of the corresponding fragmentation patterns. Apart from the C12 acyl-correlated fragments of 397.3, 515.6 and 597.6, the "+200" fragments corresponding to 515.6 and 597.6 of C12 acyl CFO (715.7 and 797.6, respectively) were observed for C12 acyl-DFO+200+30 (**Figure 4.7**). Similarly, the "+200" fragments corresponding to 515.6 and 597.6 (715.7 and 797.6, respectively), as well as the "+200+200" fragments of 515.6 (915.6 and 997.7, respectively) were observed for C12 acyl-DFO+200+200+30, as expected (**Figure 4.7**, **Appendix Figures 4.5** and **4.6**). Additionally, the two common fragments 631.5 and 549.4 (underlined by dashed lines) for all members of acyl-DFOs+200+30, are the "+200" version of the common fragments 431.3 and 349.3 observed for acyl-DFOs+30. Similarly, the fragments 631.5 and 549.5, 831.6 and 749.6 common to all members of acyl-DFOs+200+200+30 correspond to the "+200" and "+200+200" version of 431.3 and 349.3, respectively (**Figure 4.7**, **Appendix Figures 4.5** and **4.6**).

4.3 Discussion

In this work, we identified a variety of desferrioxamines (DFOs) from wild type *S. ambofaciens* and select mutant strains, using mass spectrometry and GNPS networking techniques. Among the observed DFOs, the acyl-DFOs are predicted to have the same structures as the desferrioxamines previously identified from *S. coelicolor* [216]. These acyl-DFOs were detected in a previous study (unpublished) of *S. ambofaciens* interacting with *Pseudomonas fluorescens* BBc6R8 upon cultivation on 26A agar medium. In addition, a subset of the compounds were detected following cultivation of *S. ambofaciens* alone on 26A and ISP2 agar media, albeit in lower yield (B. Aigle, personal communication). Compared to the wild type *S. ambofaciens* cultured alone on 26A agar medium, acyl-DFOs+30, acyl-DFOs+200+30, and acyl-DFOs+200+200+30 are produced at particularly high levels when grown in liquid MP5 medium, reflecting the fact that nutrients and precursors present under different cultivation conditions can affect the structural diversity within a particular metabolite family.

Structurally, acyl-DFOs are desferrioxamine with a fully saturated fatty acid tail attached to one of the three hydroxamate groups. The variability of aliphatic chain length indicates a level of flexibility in the incorporation step of the fatty acid tails. In terms of their physical and chemical properties, acyl-DFOs are a suite of amphiphilic siderophores with the appended fatty acid tails conferring high membrane affinity [216]. Amphiphilic siderophores are commonly produced by marine bacteria (e.g. marinobactins [223], aquachelins [223], amphibactins [224], ochrobactins [225], and synechobactins [226]), but are rare in terrestrial bacteria. Comparative analyses have demonstrated that certain amphiphilic siderophores (with acyl-chain lengths from C14−C21) can embed themselves into the cell membrane [227], while others (with acyl-chain length from C12−C16) can result in micelle and vesicle formation [224], suggestive of an adaptation that helps to retain siderophores within the bacterial cells, thus reducing their diffusion into the open ocean environment. In this context, their relative rarity in terrestrial bacteria (and therefore their reduced need) is presumably explained by the physical structure of soil that intrinsically limits the diffusion the molecules [228]. Therefore, all acylated DFOs identified in this work from *S. ambofaciens*, a strain isolated from soil in the north of France [158], implying an unexpected commonality between soil and marine bacteria.

Notably, our work has identified three series of potentially novel desferrioxamine-related compounds in bacteria (the acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30). The MS/MS spectra and molecular networking analysis strongly argue these new classes of DFOs are related to acyl-DFOs, but with incorporation of additional HSC units (+200) and a novel, as yet unidentified modification to the amine terminal end of the molecule (+30). At present, we can propose two alternative origins for this moiety, the first of which is based on in-source (ESI) oxidation of the terminal amine to form a NO_2 group during the ionisation process (and thus these derivatives would be artefacts), and the second based on incorporation of a 1,3-diaminopropane-based building block.

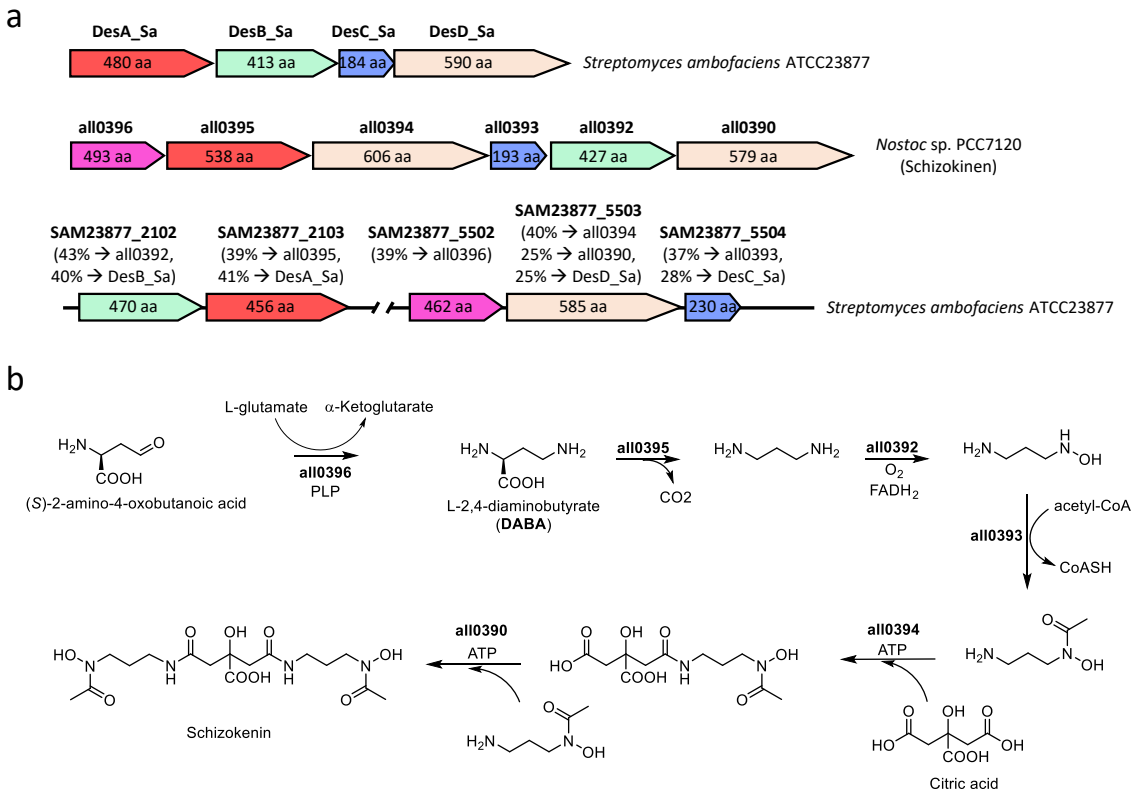
In general, the structures of NIS-mediated siderophores, especially the hydroxamate-type, contain a moiety of diamine, such as 1,3-diaminopropane, 1,4-diaminobutane (putrescine) and 1,5-diaminopentane (cadaverine) (**Appendix Figure 4.7**). The backbone origins for 1,3-diaminopropane, putrescine and cadaverine are (*S*)-2-amino-4-oxobutanoic

acid (**Figure 4.9**), L-ornithine (**Appendix Figure 4.7**) and L-lysine (**Figure 4.2**), respectively. For example, in the biosynthesis pathway to schizokenin (**Figure 4.9**), (*S*)-2-amino-4-oxobutanoic acid is transformed to 1,3-diaminopropane by sequential action of an L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase (DABA AT, all0396) and a L-2,4-diaminobutyrate decarboxylase (DABA DC, all0395). The DABA AT and DABA DC pair is also found in the BGC of the 1,3-diaminopropane-based siderophores synechobactins [226], rhizobactin 1021 [229], and acinetoferrin [230]. Interestingly, a previous study by Burrell *et al*. [231] demonstrated that although the amino sequence of DesA is more highly similar to DABA DC than lysine decarboxylase (LDC), the DABA DC homologue (DesA) preferentially decarboxylate L-lysine rather than DABA, due to the absence of a DABA AT-encoding ORF in *S. coelicolor* A3(2).

Interestingly, we identified two additional loci on the chromosome of *S. ambofaciens*, one of which contains homologs of all0392 (43% identity at the AA sequence level) and all0395 (39%) involved in the biosynthesis of schizokenin, while the other comprises homologs of all0396 (39%), all0393 (37%), and all0394 (40%)/or all0390 (25%) (**Figure 4.9a**). Nonetheless, the loci do not contain homologues to both NIS enzymes required for schizokenin biosynthesis (all0394 and all0390), as only a homologue to all0394 (SAM23877_5503) was identified. This observation likely explains the failure to detect schizokenin from cultures of *S. ambofaciens* wild type and its mutants. In view of these observations and assuming that the +30 corresponds to incorporation of 1,3-diaminopropane-derived building block (**Figure 4.8**), we proposed that the three new series of DFOs (acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30) in *S. ambofaciens* are synthesized collaboratively by enzymes encoded by the desferrioxamine and schizokenin-like gene clusters.

In this scenario, DesABC would be responsible for generation of the cadaverine-based units HSC and *N*-hydroxy-*N*-acyl-cadaverine, while the enzyme pair SAM23877_5502 (DABA AT)/SAM23877_2103 (DABA DC) or SAM23877_5502/DesA_Sa would be responsible for producing 1,3-diaminopropane. Subsequently, 1,3-diaminopropane would be *N*-hydroxylated by DesB and SAM23877_2102 (or *N*-hydroxylated twice by one of the two enzymes) on its two amino termini, followed by *N*-acetylation and *N*-succinylation by DesC and SAM23877_5504 (or by one of the two), respectively. Finally, all monomers are

linked together by amide bonds by the two NIS synthetases DesD and SAM23877_5503 together (or alternatively one of the two), resulting in the synthesis of acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30.



c

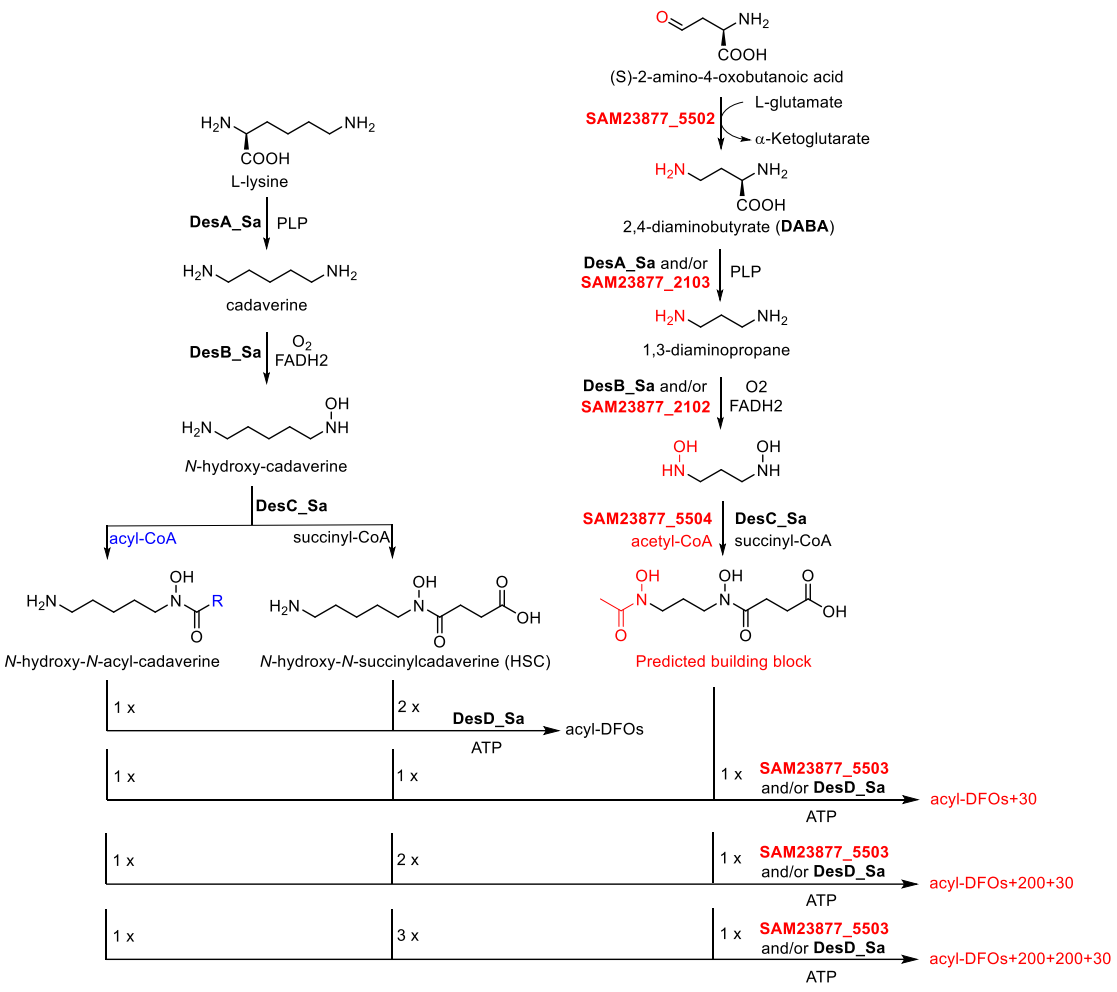


Figure 4.9 Predicted biosynthetic pathway to all DFO variants described in this study. a) The genomic organization of the potential gene cluster involved in the biosynthesis of new constituent unit, relative to the enzymes involved in the biosynthesis of schizokenin in *Nostoc* sp. PCC7120 and the DesABCD in *S. ambofaciens*. Identities on the AA sequence level are shown. Enzyme all0391 (MFS transporter) and DesE and DesF are not shown. **b**) Proposed biosynthetic pathway to schizokenin. **c**) Predicted biosynthetic pathway to the acyl-DFOs+30, acyl-DFOs+200+30 and acyl-DFOs+200+200+30.

Notably, the predicted structures of acyl-DFOs+200+30 and acyl-DFOs+200+200+30 contains four and five hydroxamic acid moieties, respectively. Recently, a chelator containing four hydroxamates, and thus which resembles DFOs+200+30, was prepared by adding an additional hydroxamic acid unit to DFO-B [232, 233]. This new compound was shown to form stable octa-coordinated complex with Zr^{4+}, and thus holds great promise for immuno-PET (positron emission tomography) imaging. This observation suggests that the

novel acyl-DFOs identified in our work might also have applications in PET. In this context, an important future goal will be to purify one or more of these compounds and elucidate their structures using NMR. The resulting compounds will be evaluated for their metal chelating activity (e.g. Fe^{3+}, Al^{3+}, Zr^{4+}), as well as their biological activities. With the structures in hand, it will then be possible to (in)validate the biosynthetic hypotheses, and in particular the interesting, and relatively rare proposed crosstalk between the two NIS gene clusters. This could be carried out, for example by inactivation of genes encoding core enzymes, such as SAM23877_2102 and 2103. Overall, this portion of the thesis illustrates that serendipitous observations can lead to the discovery of novel biosynthetic pathways and associated enzymology.

4.4 Materials and Methods

4.4.1 General

Strains used in this chapter include the mutasynthesis mutant ATCC/OE484/Δ483 (**Chapter II**), the docking domain engineering mutants K7N1–3/OE484, CPN1–2/OE484 (**Chapter III**), the wide type ATCC/OE484 (**Chapter III**), and the mutant ATCC/OE484/Δ467 constructed previously [40]. Fermentation medium and cultivation conditions are the same as described in **Chapter II**. The HPLC-MS analysis of fermentation extracts was conducted on: i) a Dionex Ultimate 3000 HPLC system coupled to a Thermo ScientificTM Q ExactiveTM Hybrid Quadrupole-Orbitrap mass spectrometer. Chromatographic separation was obtained using a Phenomenex Kinetex 2.6 μm XB-C18 150 × 4.6 mm column with solvents (A, H_2O + 0.1% formic acid) and (B, MeCN + 0.1% formic acid) using the following gradient: flow rate 0.7 mL min^{-1}, 20% B for 2 min, 20% to 98% B over 18 min, 98% B for 5 min, 98% to 20% B over 0.5 min and 20% B for 4 min (The data presented in **Figure 4.3** was collected by this machine); and ii) a Dionex Ultimate 3000 HPLC system coupled to a Thermo Scientific Orbitrap LTQ XL mass spectrometer. Chromatographic separation was obtained using the same column and flow program described in **Chapters II** and **III**.

4.4.2 Generation of the molecular network

The raw data obtained from the LC-MS/MS system were converted to mzXML format using the ProteoWizard tool msconvert. All mzXML data were uploaded to the Global Natural Products Social Molecular Networking-platform (GNPS) [217] and analyzed using the workflow published by Wang and co-workers [217], generating a molecular network based on the fragmentation spectra. To create consensus spectra, data were clustered by applying a parent mass tolerance of 1.0 Da and an MS/MS fragment ion tolerance of 0.5 Da; the remaining parameters were set to default. The output of the molecular networks was visualized using Cytoscape [234] and displayed using the settings "preferred layout" with "directed" style. For identified compounds with known structures, e.g., DFO-B and acyl-DFOs with GNPS compounds ID hit identified in the network, the fragments were assigned to the corresponding molecules by comparison to the published fragmentation patterns. The structure prediction of compounds not identified by GNPS as hits but linked to known compounds, was based on comparison of fragment spectra. This allowed identification of both common and unique fragments, and thus the structural differences between the compounds.

Chapter V Conclusion and perspectives

The work described in this thesis, carried out over the last 3.5 years, aimed to exploit key insights into modular polyketide biosynthesis to further efforts using synthetic biology to obtain interesting polyketide analogues for biological evaluation. The chosen model system, the stambomycin polyketide synthase (PKS), offered several attractive features in this regard: an unusually large size, making it an attractive target for truncation to yield smaller derivatives, and intrinsic substrate tolerance (notably that of the acyl transferase domain of module 12 (AT_{12}) and all of the downstream modules) which results in the generation of a small family of related metabolites.

Chapter II (and the resulting manuscript under evaluation by *Nat. Commun.*) reported a comprehensive suite of experiments which were deployed to generate ring-contracted stambomycins. Although a strategy based on the 'alternative module definition' (i.e. preserving intact the key ACP_n/KS_{n+1} interface at the newly-installed junction), was ultimately successful, the low yields of the obtained analogues show that even a multi-faceted approach based on a large volume of literature data, remains inefficient. Thus, the goal of efficient PKS synthetic biology continues to be distant. Nonetheless, the work reported here highlighted several additional features of hybrid systems that can be targeted in future to improve their function, including the substrate specificity of domains/modules downstream of engineered interfaces. These types of experiments could be coupled with large-scale bioinformatics analysis to better understand the natural patterns of evolution of *cis*-AT PKSs (as carried out recently for *trans*-AT PKSs [235]), which might aid in better identifying functional fusion strategies, but additionally by deploying directed evolution methods [128] to overcome the limits of rational manipulation of these systems.

Chapter III described our successful attempts to leverage the native substrate promiscuity of AT_{12} via a mutasynthesis approach. By inactivation of one of two enzymes mediating the biosynthesis of the six native extender units, we demonstrated successful generation of 6 novel stambomycins modified at the C-26 position, via incorporation of variant extender units derived from commercially-available diacids. Activation to their corresponding CoA thioesters depended on the intrinsic tolerance of the native CoA synthetase (MatB_Sa), but we additionally showed that introduction of a homologue from

S. cinnamonensis further improved the incorporation of certain analogues. Thus, this tool kit of a broad-specificity AT and MatB enzymes may be useful for generating similarly modified derivatives of additional polyketides (e.g. via an AT or module swapping approach). Furthermore, as the functionality at the C-16 position has a direct impact on the biological activity of the parental stambomycins [40], the new compounds are attractive targets for evaluation against available cancer cell lines. A declaration of invention has already been filed on these compounds with the SATT Sayens, as a prelude to obtaining patent protection. Efforts to purify, structurally characterize and biologically evaluate the stambomycin analogues are already underway (collaboration J Piel, ETH Zurich, CH).

Finally, this PhD research resulted in the surprising observation of a series of metabolites unrelated to stambomycins produced by *S. ambofaciens* wild type and several of the engineered mutant strains. By careful analysis based on a molecular network generated by the Global Natural Products Social Molecular Networking-platform (GNPS), we determined that these compounds are likely unprecedented members of the desferrioxamine (DFO) siderophore family of bacterial ion chelators. Our current hypothesis is that these derivatives result from incorporation of additional units of the classical building block HSC, but putatively also a monomer derived from the (likely inactive) schizokenin pathway present in *S. ambofaciens*, 1,3-diaminopropane, a rare example of interpathway cross-talk in siderophore biosynthesis [236]. We are currently aiming to validate these structural hypotheses by purifying and structurally characterizing the novel metabolites (collaboration J Piel, ETH Zurich, CH), as well as by inactivating specific genes in the schizokenin pathway to confirm their implication. In view of the fact that extended DFOs have recently shown to be capable of chelating PET-compatible metals (such as Zr^{4+}) [232, 233], it will also be of interest to evaluate the metal chelating properties of the novel DFOs identified in this work.

Reference

1. de Kraker, M.E., A.J. Stewardson, and S. Harbarth, *Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?* PLoS Med, 2016. **13**(11): p. e1002184.
2. Park, N.H., et al., *Addressing Drug Resistance in Cancer with Macromolecular Chemotherapeutic Agents.* J Am Chem Soc, 2018. **140**(12): p. 4244-4252.
3. Stewart BW, W.C., *IARC World cancer report.* 2014.
4. Staunton, J. and B. Wilkinson, *Biosynthesis of Erythromycin and Rapamycin.* Chem Rev, 1997. **97**(7): p. 2611-2630.
5. Abou-Zeid, A.Z., A.e.-G. Khalil, and M. Rabei, *Spiramycin, a macrolide antibiotic.* Zentralbl Bakteriol Naturwiss, 1980. **135**(5): p. 443-53.
6. Bollag, D.M., et al., *Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action.* Cancer Res, 1995. **55**(11): p. 2325-33.
7. Martin, G.D., et al., *Marmycins A and B, cytotoxic pentacyclic C-glycosides from a marine sediment-derived actinomycete related to the genus Streptomyces.* J Nat Prod, 2007. **70**(9): p. 1406-9.
8. Demain, A.L., *From natural products discovery to commercialization: a success story.* Journal of Industrial Microbiology and Biotechnology, 2006. **33**(7): p. 486-495.
9. Bérdy, J., *Bioactive microbial metabolites.* J Antibiot (Tokyo), 2005. **58**(1): p. 1-26.
10. Ligon, B.L., *Penicillin: its discovery and early development.* Semin Pediatr Infect Dis, 2004. **15**(1): p. 52-7.
11. Demain, A.L. and J.L. Adrio, *Contributions of Microorganisms to Industrial Biology.* Molecular Biotechnology, 2007. **38**(1): p. 41.
12. Demain, A.L., *Importance of microbial natural products and the need to revitalize their discovery.* Journal of Industrial Microbiology & Biotechnology, 2014. **41**(2): p. 185-201.
13. Corre, C. and G.L. Challis, *New natural product biosynthetic chemistry discovered by genome mining.* Nat Prod Rep, 2009. **26**(8): p. 977-86.
14. Velásquez, J.E. and W.A. van der Donk, *Genome mining for ribosomally synthesized natural products.* Curr Opin Chem Biol, 2011. **15**(1): p. 11-21.
15. Scheffler, R.J., et al., *Antimicrobials, drug discovery, and genome mining.* Appl Microbiol Biotechnol, 2013. **97**(3): p. 969-78.
16. Bachmann, B.O., S.G. Van Lanen, and R.H. Baltz, *Microbial genome mining for accelerated natural products discovery: is a renaissance in the making?* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 175-84.
17. Ziemert, N., M. Alanjary, and T. Weber, *The evolution of genome mining in microbes - a review.* Nat Prod Rep, 2016. **33**(8): p. 988-1005.
18. Chu, L., et al., *Genome mining as a biotechnological tool for the discovery of novel marine natural products.* Crit Rev Biotechnol, 2020. **40**(5): p. 571-589.
19. Kenshole, E., et al., *Natural product discovery through microbial genome mining.* Curr Opin Chem Biol, 2020. **60**: p. 47-54.
20. Barka, E.A., et al., *Taxonomy, Physiology, and Natural Products of Actinobacteria.* Microbiology and Molecular Biology Reviews, 2016. **80**(1): p. 1.
21. Chater, K.F., et al., *The complex extracellular biology of Streptomyces.* FEMS Microbiol Rev, 2010. **34**(2): p. 171-98.
22. Schlimpert, S., K. Flärdh, and J. Buttner, *Fluorescence Time-lapse Imaging of the Complete S. venezuelae Life Cycle Using a Microfluidic Device.* J Vis Exp, 2016(108): p. 53863.
23. Manivasagan, P., et al., *Pharmaceutically active secondary metabolites of marine actinobacteria.* Microbiol Res, 2014. **169**(4): p. 262-78.

24. Bentley, S.D., et al., *Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).* Nature, 2002. **417**(6885): p. 141-147.
25. Gomez-Escribano, J.P. and M.J. Bibb, *Streptomyces coelicolor as an expression host for heterologous gene clusters.* Methods Enzymol, 2012. **517**: p. 279-300.
26. Bérdy, J., *Thoughts and facts about antibiotics: Where we are now and where we are heading.* The Journal of Antibiotics, 2012. **65**(8): p. 385-395.
27. Schiewe, H.J. and A. Zeeck, *Cineromycins, gamma-butyrolactones and ansamycins by analysis of the secondary metabolite pattern created by a single strain of Streptomyces.* J Antibiot (Tokyo), 1999. **52**(7): p. 635-42.
28. Pan, R., et al., *Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review.* Front Microbiol, 2019. **10**: p. 294.
29. Demain, A.L. and A. Fang, *The natural functions of secondary metabolites.* Adv Biochem Eng Biotechnol, 2000. **69**: p. 1-39.
30. Sattler, I., R. Thiericke, and A. Zeeck, *The manumycin-group metabolites.* Natural Product Reports, 1998. **15**(3): p. 221-240.
31. König, C.C., et al., *Bacterium Induces Cryptic Meroterpenoid Pathway in the Pathogenic Fungus Aspergillus fumigatus.* ChemBioChem, 2013. **14**(8): p. 938-942.
32. Bertrand, S., et al., *Metabolite induction via microorganism co-culture: A potential way to enhance chemical diversity for drug discovery.* Biotechnology Advances, 2014. **32**(6): p. 1180-1204.
33. De Roy, K., et al., *Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities.* Environmental Microbiology, 2014. **16**(6): p. 1472-1481.
34. Marmann, A., et al., *Co-Cultivation-A Powerful Emerging Tool for Enhancing the Chemical Diversity of Microorganisms.* Marine Drugs, 2014. **12**(2): p. 1043-1065.
35. Pham, J.V., et al., *A Review of the Microbial Production of Bioactive Natural Products and Biologics.* Frontiers in Microbiology, 2019. **10**: p. 1404.
36. Weber, T., et al., *antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters.* Nucleic Acids Research, 2015. **43**(W1): p. W237-W243.
37. Medema, M.H., et al., *Minimum Information about a Biosynthetic Gene cluster.* Nature Chemical Biology, 2015. **11**(9): p. 625-631.
38. Bösch, N.M., et al., *Landornamides: Antiviral Ornithine-Containing Ribosomal Peptides Discovered through Genome Mining.* Angew Chem Int Ed Engl, 2020. **59**(29): p. 11763-11768.
39. Skinnider, M.A., et al., *Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences.* Nat Commun, 2020. **11**(1): p. 6058.
40. Laureti, L., et al., *Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(15): p. 6258-6263.
41. Li, S., et al., *Activating a Cryptic Ansamycin Biosynthetic Gene Cluster To Produce Three New Naphthalenic Octaketide Ansamycins with n-Pentyl and n-Butyl Side Chains.* Organic Letters, 2015. **17**(15): p. 3706-3709.
42. Pickens, L.B., Y. Tang, and Y.-H. Chooi, *Metabolic Engineering for the Production of Natural Products.* Annual Review of Chemical and Biomolecular Engineering, 2011. **2**(1): p. 211-236.
43. Li, S.M. and L. Heide, *New Aminocoumarin Antibiotics from Genetically Engineered Streptomyces Strains.* Current Medicinal Chemistry, 2005. **12**(4): p. 419-427.
44. Weissman, K.J., *Mutasynthesis – uniting chemistry and genetics for drug discovery.* Trends in Biotechnology, 2007. **25**(4): p. 139-142.
45. Ohagan, D., *BIOSYNTHESIS OF FATTY-ACID AND POLYKETIDE METABOLITES.* Natural Product Reports, 1995. **12**(1): p. 1-32.

46. Torres, J.P. and E.W. Schmidt, *The biosynthetic diversity of the animal world.* J Biol Chem, 2019. **294**(46): p. 17684-17692.
47. Torres, J.P., et al., *Animal biosynthesis of complex polyketides in a photosynthetic partnership.* Nat Commun, 2020. **11**(1): p. 2882.
48. Weissman, K.J., *Introduction to polyketide biosynthesis.* Methods Enzymol, 2009. **459**: p. 3-16.
49. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Natural Product Reports, 2001. **18**(4): p. 380-416.
50. Weissman, K.J. and P.F. Leadlay, *Combinatorial biosynthesis of reduced polyketides.* Nature Reviews Microbiology, 2005. **3**(12): p. 925-936.
51. Wagner, D.T., et al., *α-Methylation follows condensation in the gephyronic acid modular polyketide synthase.* Chemical Communications, 2016. **52**(57): p. 8822-8825.
52. Keatinge-Clay, A.T., *Stereocontrol within polyketide assembly lines.* Natural Product Reports, 2016. **33**(2): p. 141-149.
53. Weissman, K.J., *Polyketide biosynthesis: understanding and exploiting modularity.* Philos Trans A Math Phys Eng Sci, 2004. **362**(1825): p. 2671-90.
54. Staunton, J. and B. Wilkinson, *Biosynthesis of Erythromycin and Rapamycin.* Chemical Reviews, 1997. **97**(7): p. 2611-2630.
55. Broadhurst, R.W., et al., *The structure of docking domains in modular polyketide synthases.* Chemistry & Biology, 2003. **10**(8): p. 723-731.
56. Awodi, U.R., et al., *Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis.* Chem Sci, 2017. **8**(1): p. 411-415.
57. Masschelein, J., et al., *A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis.* Nat Chem, 2019. **11**(10): p. 906-912.
58. Nguyen, T., et al., *Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection.* Nature Biotechnology, 2008. **26**(2): p. 225-233.
59. Alt, S. and B. Wilkinson, *Biosynthesis of the Novel Macrolide Antibiotic Anthracimycin.* ACS Chem Biol, 2015. **10**(11): p. 2468-79.
60. Chen, X.H., et al., *Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42.* J Bacteriol, 2006. **188**(11): p. 4024-36.
61. Simunovic, V., et al., *Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases.* Chembiochem, 2006. **7**(8): p. 1206-20.
62. Dorival, J., et al., *Insights into a dual function amide oxidase/macrocyclase from lankacidin biosynthesis.* Nat Commun, 2018. **9**(1): p. 3998.
63. Weissman, K.J., *Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology.* Natural Product Reports, 2016. **33**(2): p. 203-230.
64. Shen, B., *Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms.* Current Opinion in Chemical Biology, 2003. **7**(2): p. 285-295.
65. Ames, B.D., et al., *Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis.* Proc Natl Acad Sci U S A, 2012. **109**(28): p. 11144-9.
66. Busch, B. and C. Hertweck, *Evolution of metabolic diversity in polyketide-derived pyrones: using the non-colinear aureothin assembly line as a model system.* Phytochemistry, 2009. **70**(15-16): p. 1833-40.
67. Gaisser, S., et al., *Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tü57.* J Bacteriol, 1997. **179**(20): p. 6271-8.

68. He, J. and C. Hertweck, *Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster.* Chem Biol, 2003. **10**(12): p. 1225-32.
69. Busch, B., et al., *Multifactorial control of iteration events in a modular polyketide assembly line.* Angew Chem Int Ed Engl, 2013. **52**(20): p. 5285-9.
70. Ridley, C.P., H.Y. Lee, and C. Khosla, *Evolution of polyketide synthases in bacteria.* Proceedings of the National Academy of Sciences, 2008. **105**(12): p. 4595.
71. Bisang, C., et al., *A chain initiation factor common to both modular and aromatic polyketide synthases.* Nature, 1999. **401**(6752): p. 502-505.
72. Hertweck, C., et al., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.* Natural Product Reports, 2007. **24**(1): p. 162-190.
73. Bergmann, S., et al., *Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans.* Nature Chemical Biology, 2007. **3**(4): p. 213-217.
74. Cortés, J., et al., *Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in Saccharopolyspora erythraea.* Mol Microbiol, 2002. **44**(5): p. 1213-24.
75. Achkar, J., et al., *Biosynthesis of phloroglucinol.* J Am Chem Soc, 2005. **127**(15): p. 5332-3.
76. Austin, M.B. and J.P. Noel, *The chalcone synthase superfamily of type III polyketide synthases.* Natural Product Reports, 2003. **20**(1): p. 79-110.
77. Chen, H., et al., *Glycopeptide antibiotic biosynthesis: Enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine.* Proceedings of the National Academy of Sciences, 2001. **98**(26): p. 14901.
78. Moore, B.S. and C. Hertweck, *Biosynthesis and attachment of novel bacterial polyketide synthase starter units.* Natural Product Reports, 2002. **19**(1): p. 70-99.
79. Chan, Y.A., et al., *Biosynthesis of polyketide synthase extender units.* Natural Product Reports, 2009. **26**(1): p. 90-114.
80. Ray, L. and B.S. Moore, *Recent advances in the biosynthesis of unusual polyketide synthase substrates.* Natural Product Reports, 2016. **33**(2): p. 150-161.
81. Chang, C., et al., *Uncovering the Formation and Selection of Benzylmalonyl-CoA from the Biosynthesis of Splenocin and Enterocin Reveals a Versatile Way to Introduce Amino Acids into Polyketide Carbon Scaffolds.* Journal of the American Chemical Society, 2015. **137**(12): p. 4183-4190.
82. Huang, G.Z., L.H. Zhang, and R.G. Birch, *Multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in Xanthomonas albilineans.* Microbiology-Uk, 2001. **147**: p. 631-642.
83. Hughes, Amanda J. and A. Keatinge-Clay, *Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Func-tional Showcasing of Streptomyces coelicolor MatB.* Chemistry & Biology, 2011. **18**(2): p. 165-176.
84. Erb, T.J., et al., *Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase.* Proceedings of the National Academy of Sciences, 2009. **106**(22): p. 8871.
85. Ray, L., et al., *A crotonyl-CoA reductase-carboxylase independent pathway for assembly of unusual alkylmalonyl-CoA polyketide synthase extender units.* Nature Communications, 2016. **7**(1): p. 13609.
86. Kim, Y.S. and J.K. Lee, *EVIDENCE THAT 2 COVALENT INTERMEDIATES, PHOSPHORYL AND MALONYL ENZYMES, ARE FORMED DURING MALONYL-COENZYME-A SYNTHETASE CATALYSIS.* Journal of Biological Chemistry, 1986. **261**(35): p. 6295-6297.
87. Bravo, J.A., et al., *High Yielding Template-Directed Syntheses of [2]Rotaxanes.* European Journal of Organic Chemistry, 1998. **1998**(11): p. 2565-2571.

88. Wu, K., et al., *The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units.* Gene, 2000. **251**(1): p. 81-90.
89. Rascher, A., et al., *Cloning and characterization of a gene cluster for geldanamycin production in Streptomyces hygroscopicus NRRL 3602.* FEMS Microbiology Letters, 2003. **218**(2): p. 223-230.
90. Rascher, A., et al., *Insights into the Biosynthesis of the Benzoquinone Ansamycins Geldanamycin and Herbimycin, Obtained by Gene Sequencing and Disruption.* Applied and Environmental Microbiology, 2005. **71**(8): p. 4862.
91. Li, W.L., et al., *Utilization of the methoxymalonyl-acyl carrier protein biosynthesis locus for cloning of the tautomycin biosynthetic gene cluster from Streptomyces spiroverticillatus.* Journal of Bacteriology, 2006. **188**(11): p. 4148-4152.
92. Karray, F., et al., *Organization of the biosynthetic gene cluster for the macrolide antibiotic spiramycin in Streptomyces ambofaciens.* Microbiology-Sgm, 2007. **153**: p. 4111-4122.
93. Mantovani, S.M. and B.S. Moore, *Flavin-Linked Oxidase Catalyzes Pyrrolizine Formation of Dichloropyrrole-Containing Polyketide Extender Unit in Chlorizidine A.* Journal of the American Chemical Society, 2013. **135**(48): p. 18032-18035.
94. Donadio, S., et al., *An erythromycin analog produced by reprogramming of polyketide synthesis.* Proceedings of the National Academy of Sciences, 1993. **90**(15): p. 7119.
95. Reid, R., et al., *A Model of Structure and Catalysis for Ketoreductase Domains in Modular Polyketide Synthases.* Biochemistry, 2003. **42**(1): p. 72-79.
96. Bevitt, D.J., J. Staunton, and P.F. Leadlay, *Mutagenesis of the dehydratase active site in the erythromycin-producing polyketide synthase.* Biochemical Society Transactions, 1993. **21**(1): p. 30S-30S.
97. Stassi, D.L., et al., *Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering.* Proceedings of the National Academy of Sciences, 1998. **95**(13): p. 7305.
98. Liu, L., et al., *Biosynthesis of 2-Nor-6-deoxyerythronolide B by Rationally Designed Domain Substitution.* Journal of the American Chemical Society, 1997. **119**(43): p. 10553-10554.
99. Ruan, X., et al., *Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives.* Journal of Bacteriology, 1997. **179**(20): p. 6416.
100. McDaniel, R., et al., *Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products.* Proceedings of the National Academy of Sciences, 1999. **96**(5): p. 1846.
101. Petkovic, H., et al., *A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase.* Journal of Antibiotics, 2003. **56**(6): p. 543-551.
102. Dunn, B.J., et al., *Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases.* Biochemistry, 2014. **53**(23): p. 3796-806.
103. Haydock, S.F., et al., *Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases.* FEBS Lett, 1995. **374**(2): p. 246-8.
104. Barajas, J.F., et al., *Engineered polyketides: Synergy between protein and host level engineering.* Synthetic and Systems Biotechnology, 2017. **2**(3): p. 147-166.
105. Sundermann, U., et al., *Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase.* ACS Chemical Biology, 2013. **8**(2): p. 443-450.
106. Zhang, F., et al., *Structural Insights into the Substrate Specificity of Acyltransferases from Salinomycin Polyketide Synthase.* Biochemistry, 2019. **58**(27): p. 2978-2986.

107. Kao, C.M., et al., *Engineered biosynthesis of a triketide lactone from an incomplete modular polyketide synthase.* Journal of the American Chemical Society, 1994. **116**(25): p. 11612-11613.
108. Hughes-Thomas, Z.A., et al., *Intermediates Released from a Polyether-Producing Polyketide Synthase Provide Insight into the Mechanism of Oxidative Cyclization.* Angewandte Chemie International Edition, 2003. **42**(37): p. 4475-4478.
109. Pacey, M.S., et al., *Novel erythromycins from a recombinant Saccharopolyspora erythraea strain NRRL 2338 pIG1 - I. Fermentation, isolation and biological activity.* Journal of Antibiotics, 1998. **51**(11): p. 1029-1034.
110. Wu, N., D.E. Cane, and C. Khosla, *Quantitative Analysis of the Relative Contributions of Donor Acyl Carrier Proteins, Acceptor Ketosynthases, and Linker Regions to Intermodular Transfer of Intermediates in Hybrid Polyketide Synthases.* Biochemistry, 2002. **41**(15): p. 5056-5066.
111. Klaus, M., et al., *Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases.* Journal of Biological Chemistry, 2016. **291**(31): p. 16404-16415.
112. Whicher, J.R., et al., *Structural rearrangements of a polyketide synthase module during its catalytic cycle.* Nature, 2014. **510**(7506): p. 560-564.
113. Gokhale, R.S., et al., *Dissecting and Exploiting Intermodular Communication in Polyketide Synthases.* Science, 1999. **284**(5413): p. 482.
114. Tsuji, S.Y., D.E. Cane, and C. Khosla, *Selective Protein−Protein Interactions Direct Channeling of Intermediates between Polyketide Synthase Modules.* Biochemistry, 2001. **40**(8): p. 2326-2331.
115. Wu, N., et al., *Assessing the Balance between Protein−Protein Interactions and Enzyme−Substrate Interactions in the Channeling of Intermediates between Polyketide Synthase Modules.* Journal of the American Chemical Society, 2001. **123**(27): p. 6465-6474.
116. Yadav, G., R.S. Gokhale, and B. Mohanty, *Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases.* Journal of Molecular Biology, 2003. **328**(2): p. 335-363.
117. Buchholz, T.J., et al., *Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains.* ACS Chemical Biology, 2009. **4**(1): p. 41-52.
118. Whicher, Jonathan R., et al., *Cyanobacterial Polyketide Synthase Docking Domains: A Tool for Engineering Natural Product Biosynthesis.* Chemistry & Biology, 2013. **20**(11): p. 1340-1351.
119. Tang, L., H. Fu, and R. McDaniel, *Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases.* Chemistry & Biology, 2000. **7**(2): p. 77-84.
120. Huang, J., et al., *Gene Replacement for the Generation of Designed Novel Avermectin Derivatives with Enhanced Acaricidal and Nematicidal Activities.* Applied and Environmental Microbiology, 2015. **81**(16): p. 5326.
121. Watanabe, K., et al., *Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases.* Journal of Biological Chemistry, 2003. **278**(43): p. 42020-42026.
122. Chandran, S.S., et al., *Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains.* Chemistry & Biology, 2006. **13**(5): p. 469-474.
123. Zhang, L., et al., *Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides.* Angewandte Chemie International Edition, 2017. **56**(7): p. 1740-1745.

124. Keatinge-Clay, A.T., *Polyketide Synthase Modules Redefined.* Angew Chem Int Ed Engl, 2017. **56**(17): p. 4658-4660.
125. Fisch, K.M., et al., *Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting.* Nature Chemical Biology, 2009. **5**(7): p. 494-501.
126. Ueoka, R., et al., *Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms.* Nature Chemical Biology, 2015. **11**(9): p. 705-712.
127. Yuzawa, S., et al., *Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production.* Acs Synthetic Biology, 2017. **6**(1): p. 139-147.
128. Wlodek, A., et al., *Diversity oriented biosynthesis via accelerated evolution of modular gene clusters.* Nat Commun, 2017. **8**(1): p. 1206.
129. Zhang, M., et al., *Biosynthesis of trioxacarcin revealing a different starter unit and complex tailoring steps for type II polyketide synthase.* Chemical Science, 2015. **6**(6): p. 3440-3447.
130. Griesgraber, G., et al., *3-Keto-11,12-carbazate derivatives of 6-O-methylerythromycin A synthesis and in vitro activity.* J Antibiot (Tokyo), 1996. **49**(5): p. 465-77.
131. Tang, L. and R. McDaniel, *Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad substrate specificity.* Chem Biol, 2001. **8**(6): p. 547-55.
132. Ward, S.L., et al., *Chalcomycin biosynthesis gene cluster from Streptomyces bikiniensis: novel features of an unusual ketolide produced through expression of the chm polyketide synthase in Streptomyces fradiae.* Antimicrob Agents Chemother, 2004. **48**(12): p. 4703-12.
133. Salomon, A.R., et al., *Apoptolidin, a selective cytotoxic agent, is an inhibitor of F0F1-ATPase.* Chem Biol, 2001. **8**(1): p. 71-80.
134. Mrudulakumari Vasudevan, U. and E.Y. Lee, *Flavonoids, terpenoids, and polyketide antibiotics: Role of glycosylation and biocatalytic tactics in engineering glycosylation.* Biotechnol Adv, 2020. **41**: p. 107550.
135. Olano, C., C. Méndez, and J.A. Salas, *Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis.* Natural Product Reports, 2010. **27**(4): p. 571-616.
136. Butler, A.R., et al., *Genetic engineering of aminodeoxyhexose biosynthesis in Streptomyces fradiae.* Nature Biotechnology, 2002. **20**(7): p. 713-716.
137. Gaisser, S., et al., *Parallel pathways for oxidation of 14-membered polyketide macrolactones in Saccharopolyspora erythraea.* Molecular Microbiology, 2002. **44**(3): p. 771-781.
138. Podust, L.M., et al., *The 1.92-A structure of Streptomyces coelicolor A3(2) CYP154C1. A new monooxygenase that functionalizes macrolide ring systems.* J Biol Chem, 2003. **278**(14): p. 12214-21.
139. Rix, U., et al., *Modification of post-PKS tailoring steps through combinatorial biosynthesis.* Nat Prod Rep, 2002. **19**(5): p. 542-80.
140. Yoon, Y.J., et al., *Generation of Multiple Bioactive Macrolides by Hybrid Modular Polyketide Synthases in Streptomyces venezuelae.* Chemistry & Biology, 2002. **9**(2): p. 203-214.
141. Goodman, C.D., et al., *Chemobiosynthesis of New Antimalarial Macrolides.* Antimicrobial Agents and Chemotherapy, 2013. **57**(2): p. 907.
142. Han, S., et al., *Functional characterization of CYP107W1 from Streptomyces avermitilis and biosynthesis of macrolide oligomycin A.* Arch Biochem Biophys, 2015. **575**: p. 1-7.
143. Li, Y., et al., *Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin.* Angew Chem Int Ed Engl, 2018. **57**(20): p. 5823-5827.

144. Grote, M. and F. Schulz, *Exploring the Promiscuous Enzymatic Activation of Unnatural Polyketide Extender Units in Vitro and in Vivo for Monensin Biosynthesis.* ChemBioChem, 2019. **20**(9): p. 1183-1189.
145. Yan, Y., et al., *Multiplexing of Combinatorial Chemistry in Antimycin Biosynthesis: Expansion of Molecular Diversity and Utility.* Angewandte Chemie International Edition, 2013. **52**(47): p. 12308-12312.
146. Zhang, L., et al., *Rational Control of Polyketide Extender Units by Structure-Based Engineering of a Crotonyl-CoA Carboxylase/Reductase in Antimycin Biosynthesis.* Angewandte Chemie International Edition, 2015. **54**(45): p. 13462-13465.
147. Lechner, A., et al., *Designed Biosynthesis of 36-Methyl-FK506 by Polyketide Precursor Pathway Engineering.* ACS Synthetic Biology, 2013. **2**(7): p. 379-383.
148. Goss, R.J.M., et al., *Generating rapamycin analogues by directed biosynthesis: starter acid substrate specificity of mono-substituted cyclohexane carboxylic acids.* Organic & Biomolecular Chemistry, 2006. **4**(22): p. 4071-4073.
149. Goss, R.J.M., et al., *An Expeditious Route to Fluorinated Rapamycin Analogues by Utilising Mutasynthesis.* ChemBioChem, 2010. **11**(5): p. 698-702.
150. Ban, Y.H., et al., *Mutational biosynthesis of a FK506 analogue containing a non-natural starter unit.* Molecular BioSystems, 2013. **9**(5): p. 944-947.
151. Kim, D.H., et al., *Mutational biosynthesis of tacrolimus analogues by fkbO deletion mutant of Streptomyces sp. KCTC 11604BP.* Applied Microbiology and Biotechnology, 2013. **97**(13): p. 5881-5892.
152. Lin, Z., D. Chen, and W. Liu, *Biosynthesis-based artificial evolution of microbial natural products.* Science China Chemistry, 2016. **59**(9): p. 1175-1187.
153. Koryakina, I., et al., *Poly Specific trans-Acyltransferase Machinery Revealed via Engineered Acyl-CoA Synthetases.* ACS Chemical Biology, 2013. **8**(1): p. 200-208.
154. Musiol-Kroll, E.M., et al., *Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII.* ACS Synthetic Biology, 2017. **6**(3): p. 421-427.
155. Möller, D., et al., *Flexible enzymatic activation of artificial polyketide extender units by Streptomyces cinnamonensis into the monensin biosynthetic pathway.* Letters in Applied Microbiology, 2018. **67**(3): p. 226-234.
156. Thirumurugan, P., D. Matosiuk, and K. Jozwiak, *Click Chemistry for Drug Development and Diverse Chemical–Biology Applications.* Chemical Reviews, 2013. **113**(7): p. 4905-4979.
157. Ye, Z., et al., *Reprogramming Acyl Carrier Protein Interactions of an Acyl-CoA Promiscuous trans-Acyltransferase.* Chemistry & Biology, 2014. **21**(5): p. 636-646.
158. Pinnert-Sindico, S., *A new species of Streptomyces producing antibiotics Streptomyces ambofaciens n. sp., cultural characteristics.* Ann Inst Pasteur (Paris), 1954. **87**(6): p. 702-7.
159. Leblond, P., et al., *The unstable region of Streptomyces ambofaciens includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA.* Mol Microbiol, 1996. **19**(2): p. 261-71.
160. Laureti, L., et al., *Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in <em>Streptomyces ambofaciens</em&gt.* Proceedings of the National Academy of Sciences, 2011. **108**(15): p. 6258.
161. Aigle, B.L., FR), Challis, Gregory (Leics, GB), Laureti, Luisa (Issy Les Moulineaux, FR), Song, Lijiang (Conventry, GB), Leblond, Pierre (Flavigny-sur-Moselle, FR), *STAMBOMYCIN AND DERIVATIVES, THEIR PRODUCTION AND THEIR USE AS DRUGS*. 2012, UNIVERSITE HENRI POINCARE NANCY 1 (Nancy Cedex, FR),INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (Paris Cedex 07, FR),UNIVERSITY OF WARWICK (Coventry, GB): United States.

162. Aigle, B., et al., *Genome mining of Streptomyces ambofaciens.* Journal of Industrial Microbiology & Biotechnology, 2014. **41**(2): p. 251-263.
163. Song, L., et al., *Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis.* The Journal of Antibiotics, 2014. **67**(1): p. 71-76.
164. Heathcote, M.L., J. Staunton, and P.F. Leadlay, *Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units.* Chem Biol, 2001. **8**(2): p. 207-20.
165. Kim, B.S., et al., *Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin.* J Biol Chem, 2002. **277**(50): p. 48028-34.
166. Keatinge-Clay, A.T., *A tylosin ketoreductase reveals how chirality is determined in polyketides.* Chem Biol, 2007. **14**(8): p. 898-908.
167. Dutta, S., et al., *Structure of a modular polyketide synthase.* Nature, 2014. **510**(7506): p. 512-517.
168. Mo, S., et al., *Biosynthesis of the allylmalonyl-CoA extender unit for the FK506 polyketide synthase proceeds through a dedicated polyketide synthase and facilitates the mutasynthesis of analogues.* J Am Chem Soc, 2011. **133**(4): p. 976-85.
169. Sandy, M., et al., *Enzymatic synthesis of dilactone scaffold of antimycins.* ACS Chem Biol, 2012. **7**(12): p. 1956-61.
170. Khosla, C., et al., *Structure and mechanism of the 6-deoxyerythronolide B synthase.* Annu Rev Biochem, 2007. **76**: p. 195-221.
171. Paiva, P., et al., *Understanding the Catalytic Machinery and the Reaction Pathway of the Malonyl-Acetyl Transferase Domain of Human Fatty Acid Synthase.* ACS Catalysis, 2018. **8**(6): p. 4860-4872.
172. Hans, M., et al., *Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules.* J Am Chem Soc, 2003. **125**(18): p. 5366-74.
173. Bravo-Rodriguez, K., et al., *Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis.* Chem Biol, 2015. **22**(11): p. 1425-1430.
174. Koryakina, I., et al., *Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering.* ACS Chem Biol, 2017. **12**(1): p. 114-123.
175. Ghisalba, O., et al., *A genetic approach to the biosynthesis of the rifamycin-chromophore in Nocardia-mediterranei* . Journal of Antibiotics, 1981. **34**(1): p. 58-63.
176. Hunziker, D., et al., *Primer unit specificity in rifamycin biosynthesis principally resides in the later stages of the biosynthetic pathways.* Journal of the American Chemical Society, 1998. **120**(5): p. 1092-1093.
177. Dutton, C.J., et al., *Novel avermectins produced by mutational biosynthesis.* Journal of Antibiotics, 1991. **44**(3): p. 357-365.
178. Zhuo, Y., et al., *Synthetic biology of avermectin for production improvement and structure diversification.* Biotechnol J, 2014. **9**(3): p. 316-25.
179. Ray, L., et al., *A crotonyl-CoA reductase-carboxylase independent pathway for assembly of unusual alkylmalonyl-CoA polyketide synthase extender units*. Nature Communications, 2016. **7**.
180. Laureti, L., *Activation d'un cluster de gènes de polycétide synthase de type I chez Streptomyces ambofaciens ATCC 23877 : isolation et caractérisation d'un nouveau macrolide géant*. 2010.
181. Zhang, W., et al., *Rational engineering acyltransferase domain of modular polyketide synthase for expanding substrate specificity*. Methods Enzymol, 2019. **622**: p. 271-292.
182. Arnold, F.H., *Directed Evolution: Bringing New Chemistry to Life.* Angew Chem Int Ed Engl, 2018. **57**(16): p. 4143-4148.

183. Hanahan, D., *Studies on transformation of Escherichia coli with plasmids.* Journal of Molecular Biology, 1983. **166**(4): p. 557-580.
184. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6640-6645.
185. Macneil, D.J., et al., *Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector.* Gene, 1992. **111**(1): p. 61-68.
186. Donadio, S. and L. Katz, *Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea.* Gene, 1992. **111**(1): p. 51-60.
187. Wilkinson, C.J., et al., *Increasing the efficiency of heterologous promoters in actinomycetes.* Journal of Molecular Microbiology and Biotechnology, 2002. **4**(4): p. 417-426.
188. Gust, B., et al., *PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(4): p. 1541-1546.
189. Zelyas, N., K. Tahlan, and S.E. Jensen, *Use of the native flp gene to generate in-frame unmarked mutations in Streptomyces spp.* Gene, 2009. **443**(1-2): p. 48-54.
190. Aubry, C., J.L. Pernodet, and S. Lautru, *Modular and Integrative Vectors for Synthetic Biology Applications in Streptomyces spp.* Appl Environ Microbiol, 2019. **85**(16).
191. Andrews, S.C., A.K. Robinson, and F. Rodríguez-Quiñones, *Bacterial iron homeostasis.* FEMS Microbiology Reviews, 2003. **27**(2-3): p. 215-237.
192. Miethke, M. and M.A. Marahiel, *Siderophore-Based Iron Acquisition and Pathogen Control.* Microbiology and Molecular Biology Reviews, 2007. **71**(3): p. 413.
193. Sheldon, J.R. and D.E. Heinrichs, *Recent developments in understanding the iron acquisition strategies of gram positive pathogens.* FEMS Microbiology Reviews, 2015. **39**(4): p. 592-630.
194. Boiteau, R.M., et al., *Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean.* Proceedings of the National Academy of Sciences, 2016. **113**(50): p. 14237.
195. Gibson, F. and D.I. Magrath, *The isolation and characterization of a hydroxamic acid (aerobactin) formed by Aerobacter aerogenes 62-1.* Biochimica et Biophysica Acta (BBA) - General Subjects, 1969. **192**(2): p. 175-184.
196. Naureen, Z., et al., *Exploring the Potentials of Lysinibacillus sphaericus ZA9 for Plant Growth Promotion and Biocontrol Activities against Phytopathogenic Fungi.* Frontiers in Microbiology, 2017. **8**: p. 1477.
197. Grandchamp, G.M., L. Caro, and E.A. Shank, *Pirated Siderophores Promote Sporulation in Bacillus subtilis.* Appl Environ Microbiol, 2017. **83**(10).
198. Galet, J., et al., *Pseudomonas fluorescens pirates both ferrioxamine and ferricoelichelin siderophores from Streptomyces ambofaciens.* Appl Environ Microbiol, 2015. **81**(9): p. 3132-41.
199. Guan, L.L., K. Kanoh, and K. Kamino, *Effect of exogenous siderophores on iron uptake activity of marine bacteria under iron-limited conditions.* Appl Environ Microbiol, 2001. **67**(4): p. 1710-7.
200. Meyer, J.M., et al., *Siderophore-mediated iron uptake in fluorescent Pseudomonas: characterization of the pyoverdine-receptor binding site of three cross-reacting pyoverdines.* Arch Biochem Biophys, 2002. **397**(2): p. 179-83.
201. Butaitė, E., et al., *Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater Pseudomonas communities.* Nat Commun, 2017. **8**(1): p. 414.
202. Kramer, J., Ö. Özkaya, and R. Kümmerli, *Bacterial siderophores in community and host interactions.* Nat Rev Microbiol, 2020. **18**(3): p. 152-163.

203. Robinson, A.E., et al., *Uropathogenic enterobacteria use the yersiniabactin metallophore system to acquire nickel.* J Biol Chem, 2018. **293**(39): p. 14953-14961.
204. Perry, R.D., A.G. Bobrov, and J.D. Fetherston, *The role of transition metal transporters for iron, zinc, manganese, and copper in the pathogenesis of Yersinia pestis.* Metallomics, 2015. **7**(6): p. 965-78.
205. Carroll, C.S. and M.M. Moore, *Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases.* Crit Rev Biochem Mol Biol, 2018. **53**(4): p. 356-381.
206. Crosa, J.H. and C.T. Walsh, *Genetics and assembly line enzymology of siderophore biosynthesis in bacteria.* Microbiol Mol Biol Rev, 2002. **66**(2): p. 223-49.
207. Oves-Costales, D., N. Kadi, and G.L. Challis, *The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis.* Chem Commun (Camb), 2009(43): p. 6530-41.
208. Pollack, J.R. and J.B. Neilands, *Enterobactin, an iron transport compound from Salmonella typhimurium.* Biochem Biophys Res Commun, 1970. **38**(5): p. 989-92.
209. Barona-Gómez, F., et al., *Identification of a cluster of genes that directs desferrioxamine biosynthesis in Streptomyces coelicolor M145.* J Am Chem Soc, 2004. **126**(50): p. 16282-3.
210. Konetschny-Rapp, S., et al., *Staphyloferrin A: a structurally new siderophore from staphylococci.* Eur J Biochem, 1990. **191**(1): p. 65-74.
211. Barbeau, K., et al., *Petrobactin, a photoreactive siderophore produced by the oil-degrading marine bacterium Marinobacter hydrocarbonoclasticus.* J Am Chem Soc, 2002. **124**(3): p. 378-9.
212. Carran, C.J., et al., *Heterobactins: A new class of siderophores from Rhodococcus erythropolis IGTS8 containing both hydroxamate and catecholate donor groups.* Biometals, 2001. **14**(2): p. 119-25.
213. Challis, G.L. and J. Ravel, *Coelichelin, a new peptide siderophore encoded by the Streptomyces coelicolor genome: structure prediction from the sequence of its non-ribosomal peptide synthetase.* FEMS Microbiol Lett, 2000. **187**(2): p. 111-4.
214. Aigle, B., et al., *Genome mining of Streptomyces ambofaciens.* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 251-63.
215. Barona-Gómez, F., et al., *Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in Streptomyces coelicolor A3(2) and Streptomyces ambofaciens ATCC 23877.* Microbiology (Reading), 2006. **152**(Pt 11): p. 3355-3366.
216. Traxler, M.F., et al., *Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome.* mBio, 2013. **4**(4).
217. Wang, M., et al., *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.* Nature Biotechnology, 2016. **34**(8): p. 828-837.
218. Senges, C.H.R., et al., *The secreted metabolome of Streptomyces chartreusis and implications for bacterial chemistry.* Proc Natl Acad Sci U S A, 2018. **115**(10): p. 2490-2495.
219. Van Berkel, G.J. and V. Kertesz, *Using the electrochemistry of the electrospray ion source.* Anal Chem, 2007. **79**(15): p. 5510-20.
220. Wabner, D. and C. Grambow, *Reactive intermediates during oxindation of water lead dioxide and platinum electrodes.* Journal of Electroanalytical Chemistry and Interfacial Electrochemistry, 1985. **195**(1): p. 95-108.
221. Lee, J., M. Simurdiak, and H. Zhao, *Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation.* J Biol Chem, 2005. **280**(44): p. 36719-27.

222. Lee, J. and H. Zhao, *Mechanistic studies on the conversion of arylamines into arylnitro compounds by aminopyrrolnitrin oxygenase: identification of intermediates and kinetic studies.* Angew Chem Int Ed Engl, 2006. **45**(4): p. 622-5.
223. Martinez, J.S., et al., *Self-assembling amphiphilic siderophores from marine bacteria.* Science, 2000. **287**(5456): p. 1245-7.
224. Martinez, J.S., et al., *Structure and membrane affinity of a suite of amphiphilic siderophores produced by a marine bacterium.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3754-9.
225. Martin, J.D., et al., *Structure and membrane affinity of new amphiphilic siderophores produced by Ochrobactrum sp. SP18.* J Biol Inorg Chem, 2006. **11**(5): p. 633-41.
226. Ito, Y. and A. Butler, *Structure of synechobactins, new siderophores of the marine cyanobacterium Synechococcus sp. PCC 7002.* Limnology and Oceanography, 2005. **50**(6): p. 1918-1923.
227. Xu, G., et al., *Membrane affinity of the amphiphilic marinobactin siderophores.* J Am Chem Soc, 2002. **124**(45): p. 13408-15.
228. Kümmerli, R., et al., *Habitat structure and the evolution of diffusible siderophores in bacteria.* Ecol Lett, 2014. **17**(12): p. 1536-44.
229. Lynch, D., et al., *Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by Sinorhizobium meliloti.* J Bacteriol, 2001. **183**(8): p. 2576-85.
230. Okujo, N., et al., *Structure of acinetoferrin, a new citrate-based dihydroxamate siderophore from Acinetobacter haemolyticus.* Biometals, 1994. **7**(2): p. 170-6.
231. Burrell, M., et al., *Evolution of a novel lysine decarboxylase in siderophore biosynthesis.* Mol Microbiol, 2012. **86**(2): p. 485-99.
232. Guérard, F., et al., *Investigation of Zr(iv) and 89Zr(iv) complexation with hydroxamates: progress towards designing a better chelator than desferrioxamine B for immuno-PET imaging.* Chemical Communications, 2013. **49**(10): p. 1002-1004.
233. Patra, M., et al., *An octadentate bifunctional chelating agent for the development of stable zirconium-89 based molecular imaging probes.* Chemical Communications, 2014. **50**(78): p. 11523-11525.
234. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
235. Helfrich, E.J.N., et al., *Evolution of combinatorial diversity in trans-acyltransferase polyketide synthase assembly lines across bacteria.* Nat Commun, 2021. **12**(1): p. 1422.
236. Lazos, O., et al., *Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters.* Chem Biol, 2010. **17**(2): p. 160-73.

Appendix

Appendix to Chapter II

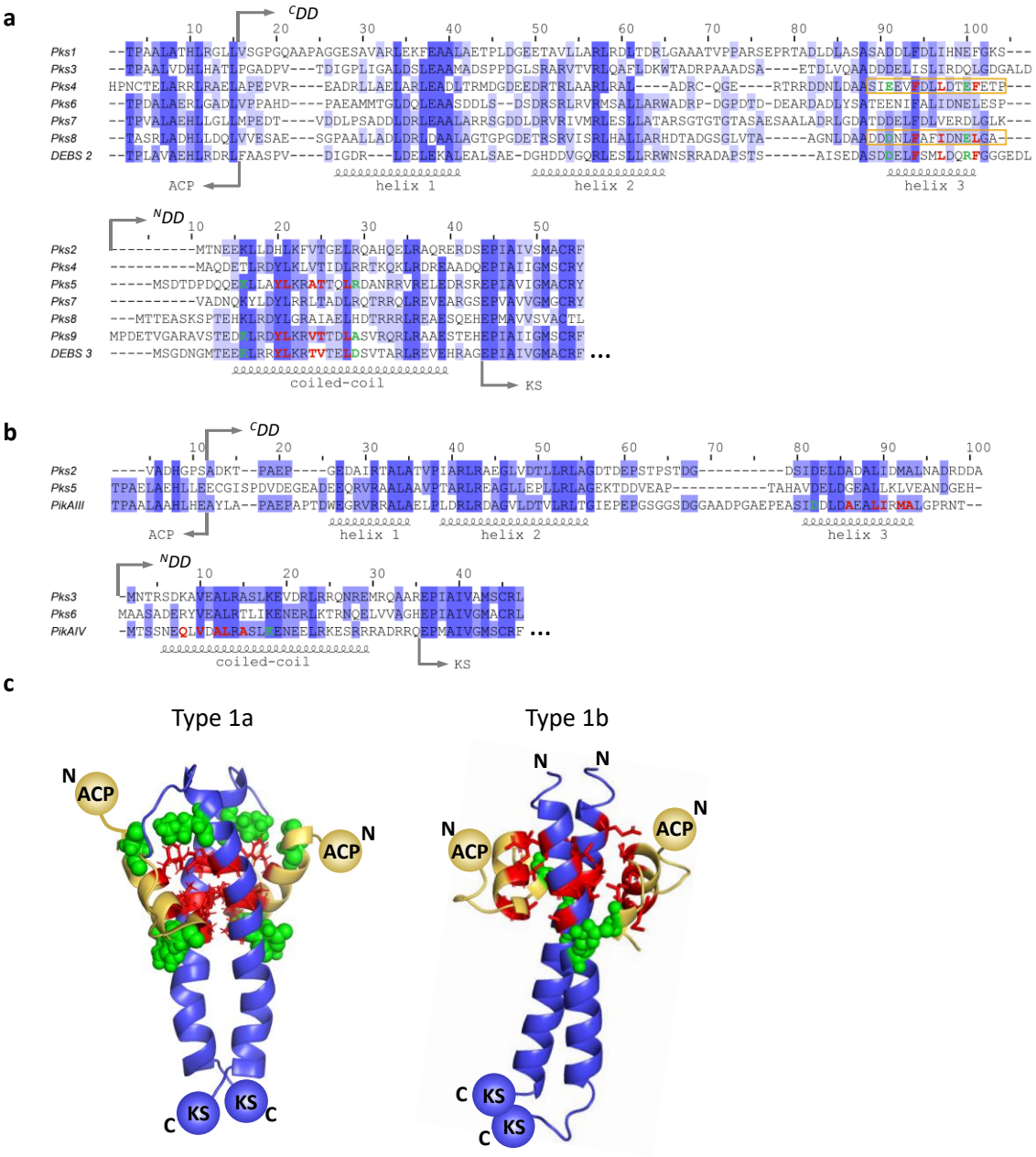
Supplementary Information

Successes, surprises and pitfalls in modular polyketide synthase engineering: generation of ring-contracted stambomycins

Li Su[1,2], Laurence Hôtel[2], Cédric Paris[3], Alexander O. Brachmann[4], Jörn Piel[4], Christophe Jacob[1]*, Bertrand Aigle[2]*, Kira J. Weissman[1]*

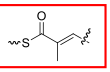
[1]Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France. [2]Université de Lorraine, INRAE, DynAMic, F-54000 Nancy, France. [3]Université de Lorraine, PASM, F-54000 Nancy, France. [4]Institute of Microbiology, ETH Zurich, 8093 Zurich, Switzerland.

Table of Contents

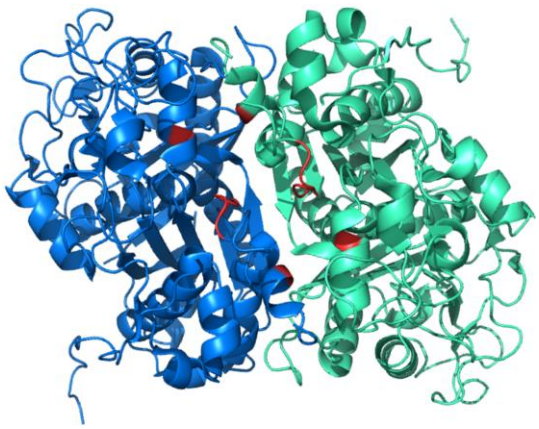


Supplementary Fig. 1 Analysis of putative DDs in the stambomycin PKS. a,b Multiple sequence alignment of stambomycin PKS docking domains. Shown are the regions of α-helical secondary structure as determined by NMR for representative type 1a (DEBS 2–3[1]; PDB: 1PZQ, 1PZR) (**a**) and type 1b (PikAIII–IV[2]; PDB: 3F5H) docking domain complexes (**b**). Conserved residues among the docking domains are highlighted in shades of blue based on the percentage identities. Red and green residues represent, respectively, the amino acids involved in the principal hydrophobic and electrostatic interactions at the docking domain interfaces. The key positions on helix α3 of $^{C}DD_4$ (E, F, L, E, F) were mutated to match those of $^{C}DD_8$ (D, F, I, E, L). The residues shown within the yellow rectangle correspond to the helix swap region. **c** Structures of the type 1a (PDB:1PZR[1]) and type 1b (PDB: 3F5H[2]) docking complexes showing the key residue positions. In each case, the ^{C}DD is represented in gold and the ^{N}DD in blue, hydrophobic residues as red sticks, and amino acids involved in electrostatic interactions as green spheres.

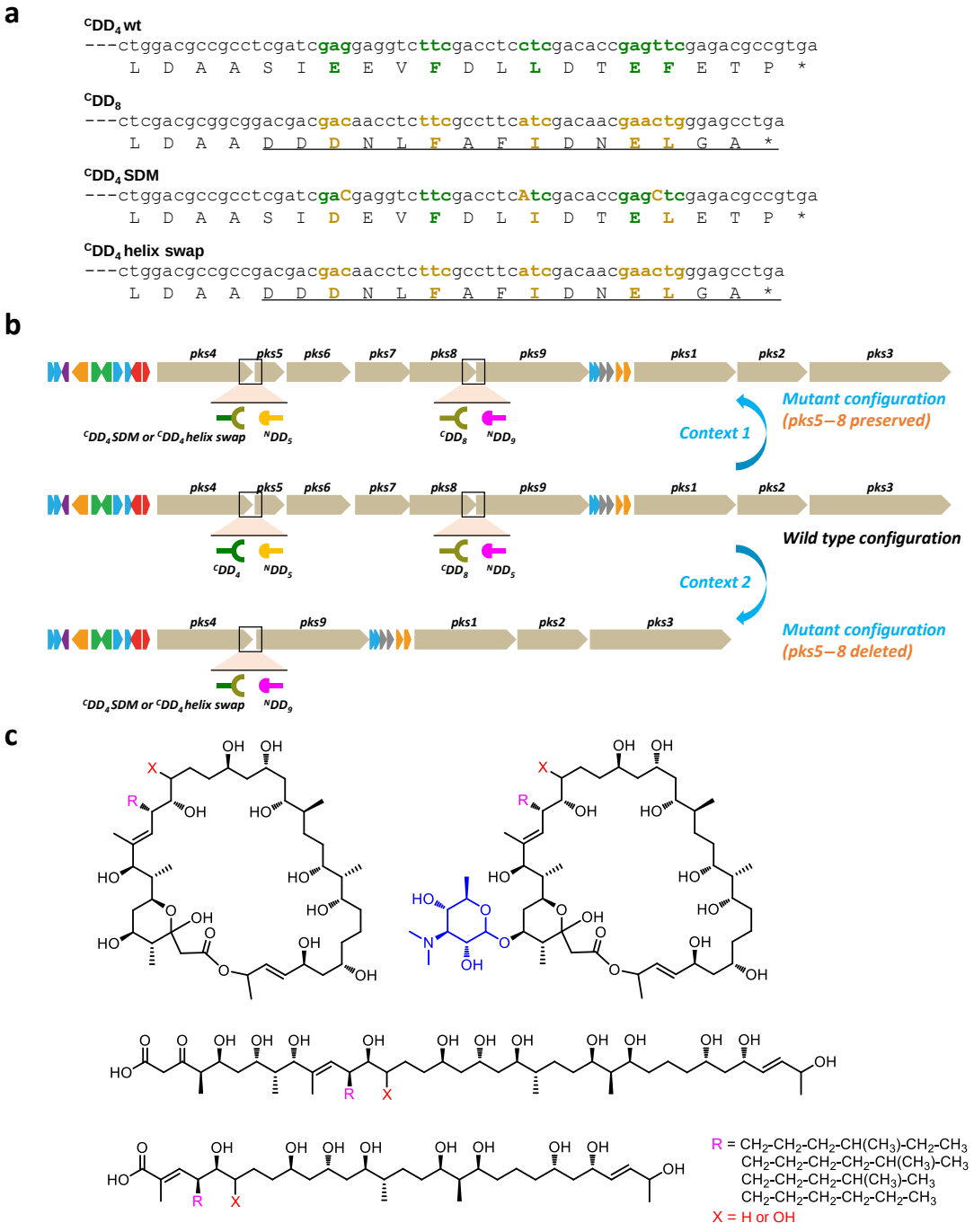
a

KS	Sequence alignment with the putative specificity motifs highlighted	Incoming substrate
DEBS KS3	-------TATGVFLG**VAKFGY**GEDTA--AAEDVEGYS**V**TGVAPAVASGRISYTM-36aa-VGGAAVMATPG**V**FVDFSRQRALAADGRSKAFGA-------	
KS3	-------SRVGVFVG**TNGQDY**AGLFPA-TGSGLEGHV**A**TGSAASVLSGRISYTY-36aa-AGGVTVMAAP**T**AFLEFAHQRGLAADGRCKAFSA-------	
KS4 (Pks1/2)	-------ARVGVFAA**TNGQDY**ATALAADPEAA-DGYL**A**TGAVASVLSGRISYVF-36aa-AGGVTVMATPA**A**FVEFSRQRGLAADGRCKAFGD-------	
KS10	-------RSVGMFVG**TNGQDY**PVVLAGSADEGLDAHAA**T**GNAAAVLSGRVSYAF-36aa-AGGVSVMSTEA**A**FTEFARQGGLADDGRCKAFSA-------	
KS15 (Pks5/6)	-------SPVGVYVG**AGTSGY**GIGVPV-AEEA-AGYA**L**TGTATSVLCGRVAYSF-36aa-AGGVTVMATPA**A**FVEFSRQRGLARDGRCKAFAA-------	
KS23	-------SDTGVFAG**TNGQDY**LALLLGSAAAT-EGHL**G**TGNSASVLSGRISYTF-36aa-AGGVTAMPLPG**T**FIEFSTQGALSSDGRCKAFSA-------	
KS6 (Pks2/3)	-------SDTGVFVG**ASPSGY**DTVG--RLPESTVGY**Q**LTGSAGSVLSGRLSYVL-36aa-VGGVAVITTPA**A**FTEFGKQGGMASDGRCKSFSA-------	
KS18	-------TPTGVFIG**ASPSGY**GVGV--AVPGS-EGHS**L**TGMSGSVLSGRVAYVF-36aa-VGGVAVMTSPG**V**FSEFARQDGLASDGRCRAFSA-------	
KS24	-------TRTGVFVG**ASAGGY**AMAG-VLPEGS-ESHA**M**TGTSNSVLSGRVSYVF-36aa-AGGVTVMHSPV**V**FAEFGRQGGLASDGRCRSFAA-------	
KS7	-------TRTGVLIG**AASSGY**GLGT--DLFTTAEGHV**L**AGGSNSVISGRVAYSF-36aa-AGGVTVMAAPS**I**FAEFNRQGGLAGDGRCKAFAG-------	
KS9	-------RPVGVFAG**ASSSGY**GGAG--DDLEGAGGYL**L**AGTANSVISGRVAYTF-36aa-AGGVTVMVSPA**A**FAEFDRQDGLASDGRCKSFAG-------	
KS11 (Pks3/4)	-------TRTGVFAG**GNDQGY**LRLLANEP--GSVGHQ**L**TGGATAVISGRVAYTL-36aa-AGGVTVMATPG**V**FTEFTRQRGLAADGRCKSFSV-------	
KS16	-------SDTAVFAG**TNGQDY**PILLAGDPDIS-EGHQ**G**AGNAAAVLSGRIAYTF-36aa-AGGVTVMSTPT**A**FVEFSRQLGLAADGRCKAFAA-------	
KS19 (Pks7/8)	-------SRTGVFIG**TTGQGY**GINMQAGMA-GTEGYQ**L**TGSATAVLSGRIAYTL-36aa-AGGVTVMPHPI**A**FHEFNRQGGLASDGRCKAFSA-------	
KS14 (Pks4/5)	-------GRTGVFAG**VMYHDY**PSVV--DPEAL-DGYL**G**TANAGSVLSGRIAYTF-36aa-VGGVTVMSGPM**M**FAGFGLEDGSAADGRCKAFAA-------	
KS21 (Pks8/9)	-------SRTGVFTG**VMYNEY**ASRIMSVPD-EVAGHL**G**TGNSGSVVSGRVAYTL-36aa-AGGVAIMVTPN**T**FFDFALSGGLASDGRCKAFSA-------	
KS2	-------SDTGVFAG**VMYHDY**AHAAEALPE--TEGYR**A**TGGAGSVVSGRIAYTY-36aa-VGGVTVMSAPG**V**FVDFSKQRGLAPDGRCKSFSA-------	
KS20	-------SSTGVFAG**LMYHDY**GATVTVLPE-GVEGYL**G**TGTAGSVLAGRVSYTF-36aa-AGGVTVLSTPG**V**FVDFSRQRGLASDGRCRSFSA-------	
KS5	-------EPIGVFAG**AASSYY**GLDK--ALPEDVAGYQ**L**TGGATSVLSGRVAYSF-36aa-AGGVTVMAGFG**V**FSEFSRLDGLAGDARCKAFAE-------	
KS12	-------SSTGVYVG**TATSGY**GLGRFAVPD-GSRPHV**L**TGTATSVVSGRLAYTF-36aa-AGGATVMAVPG**I**FTDAAQGGALAPGGRCRAFSS-------	
KS17 (Pks6/7)	-------SRTGVFIG**SSSSAY**GTGLRTLPQ-GVEGHL**L**TGSAPSVVSGRVAYAL-36aa-AGGVTVMTSPG**I**FTEFSKQRGLAADGRCKPFSA-------	
KS1	-------SRTGVFVG**ATAQDY**GPRMHE-PAESSEGYL**L**TGTTASVASGRIAYAF-36aa-AGGVTLMATPG**V**LVEFARQDGLSPDGRCKAFGA-------	
KS8	-------SSTGVFIG**AAHAQY**GYGM--RLPDNALGHL**M**TGTTTSVASGRLAYTF-36aa-AGGVTVLATPG**V**ITEFDRQRGLASDGRCKAFSA-------	
KS13	-------SPTGVFVG**TSFVGY**GIGAQQ-PGNEAEGFP**L**AGTGTAAASGRISYTF-36aa-AGGAAVLATPA**S**FTEFSRQRGLAADGRCKPFAA-------	
KS22	-------SRTAVYAG**TGGODY**LAVLAGDPVAG-EGYL**V**TGGSPSVLSGRVAYAF-36aa-VGGVNVISLPS**V**FAEFSKQRGQAADGRCKAFSA-------	

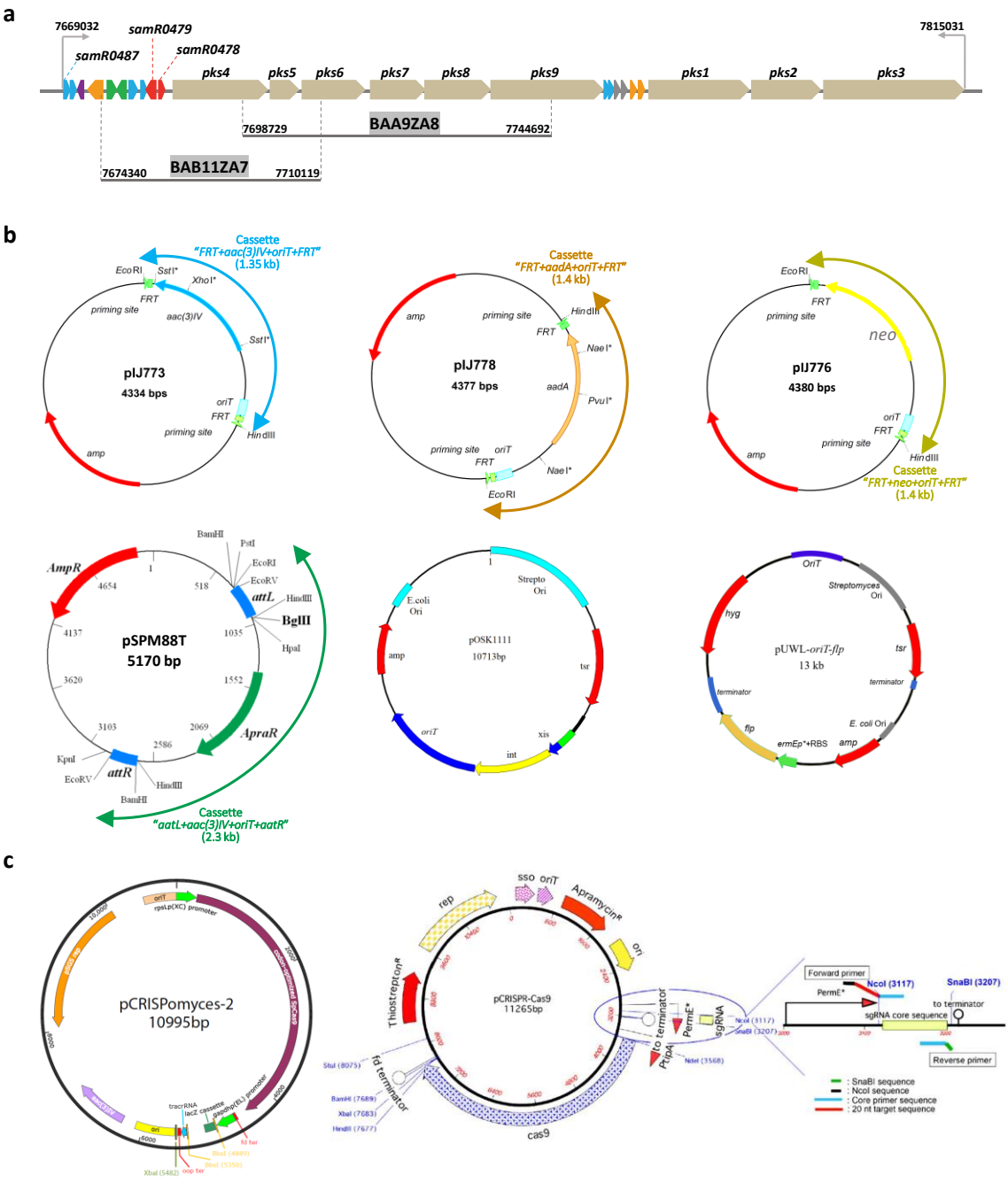
b



Supplementary Fig. 2 Analysis of the incoming substrates and the putative specificity motifs of the stambomycin PKS KS domains. a Sequence alignment showing the residues in the stambomycin PKS KS domains corresponding to those identified in erythromycin PKS (DEBS) KS_3 as mediating substrate specificity[3]. In this figure, stambomycin KS domains which accept substrates of identical structure at the α- and β-positions have been grouped together. The brackets indicate intersubunit junctions where the corresponding KS domains are present, e.g. for KS_4 (Pks1/2). **b** Locations within the KS domains of the putative substrate specificity motifs (VAKFGY-V-V, shown in red within the dimer structure of DEBS KS_3–AT_3 didomain[3] (PDB: 2QO3, from which the AT domains have been omitted).

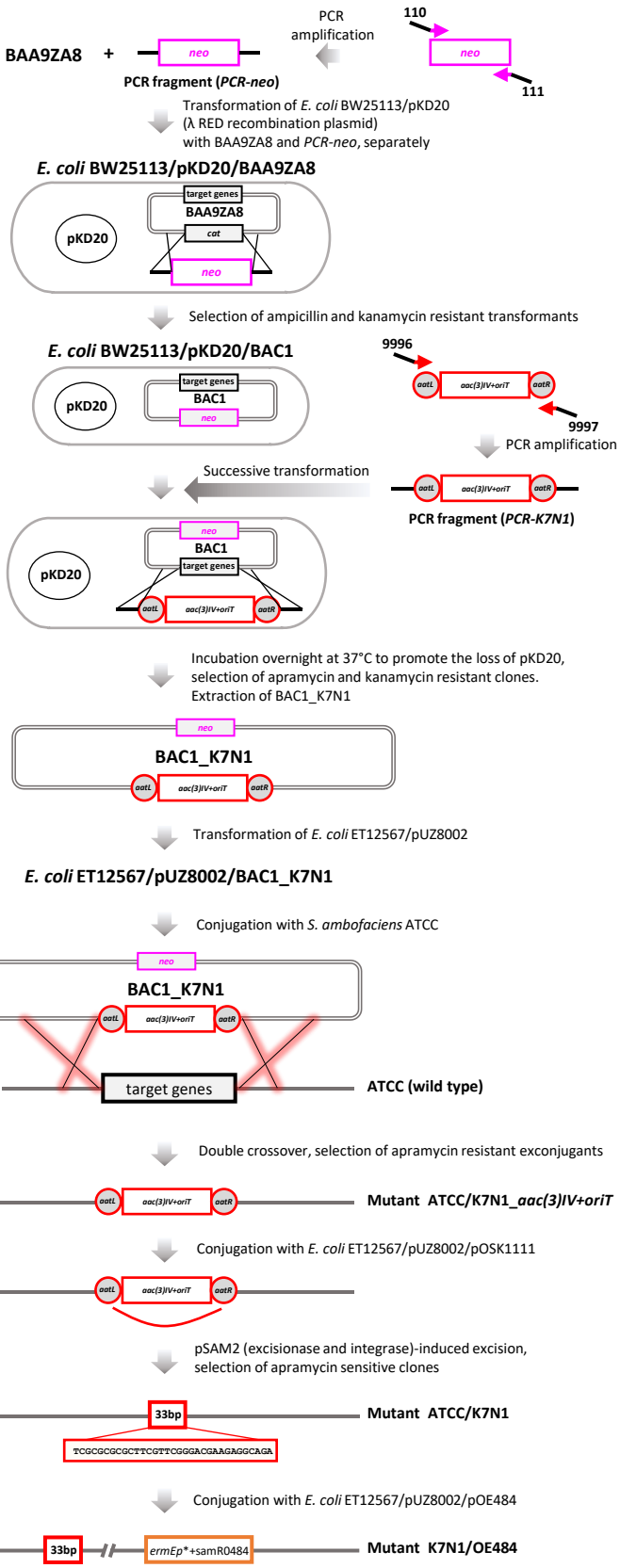


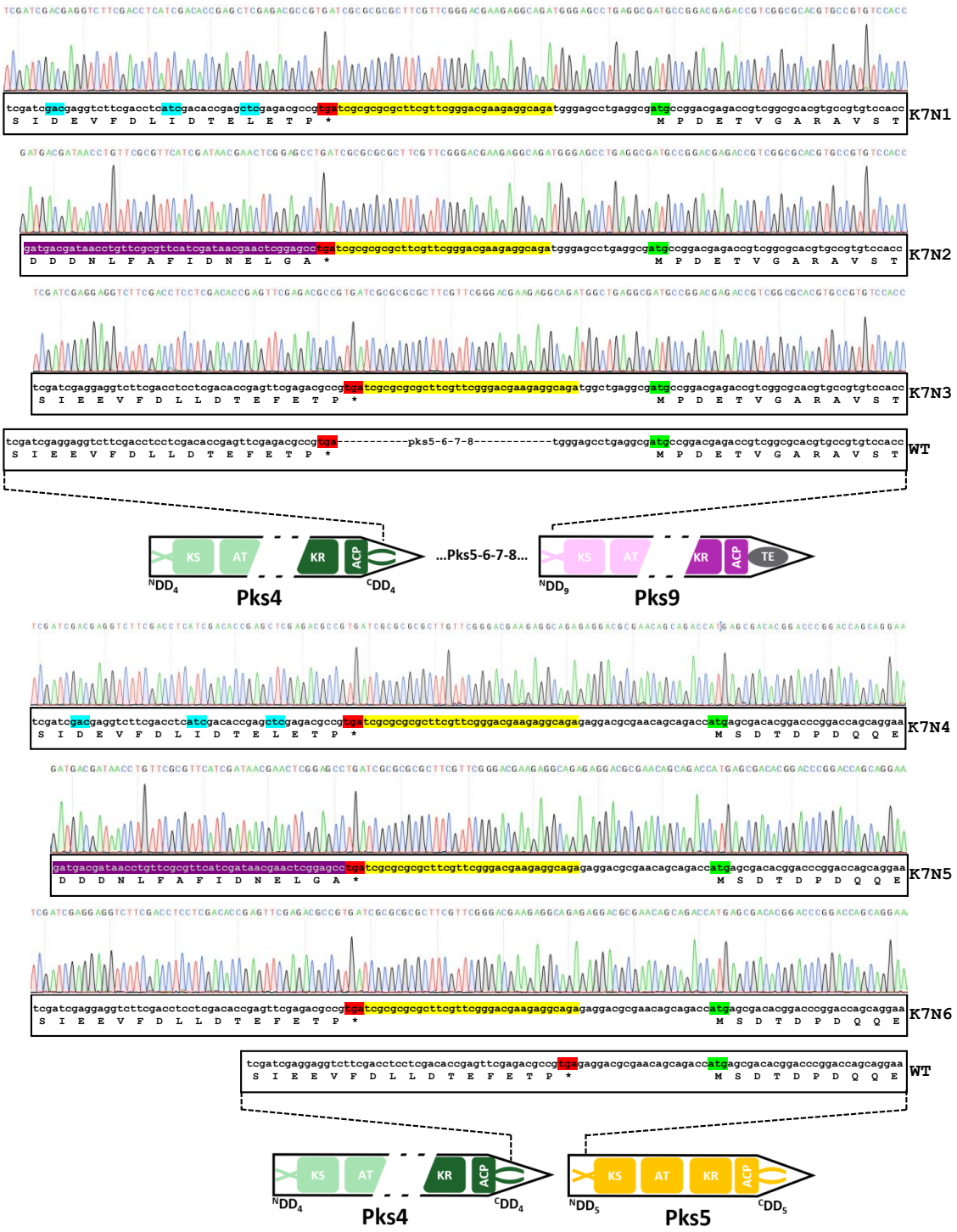
Supplementary Fig. 3 Overview of the interface engineering strategy and the predicted metabolites. a DD modification strategy. Potentially critical residues mediating interaction specificity are colored in dark green in $^{C}DD_4$ wt and orange in $^{C}DD_8$, respectively. As indicated, $^{C}DD_4$ SDM was generated by site-directed mutagenesis (SDM) within $^{C}DD_4$, and $^{C}DD_4$ helix swap by exchange of the full helix α3. **b** Two genetic contexts for the modifications. In the first, the intervening subunits (Pks5–Pks8) were deleted in order to remove competition from subunit 5 for subunit 4, and subunit 8 for subunit 9, while in the second, the intervening genes were preserved. **c** A selection of metabolites potentially generated by a PKS incorporating a novel interface between Pks4 and Pks9. Position X corresponds to the site of potential hydroxylation by cytochrome P450 SamR0478. All indicated stereochemistries are based on prediction only.



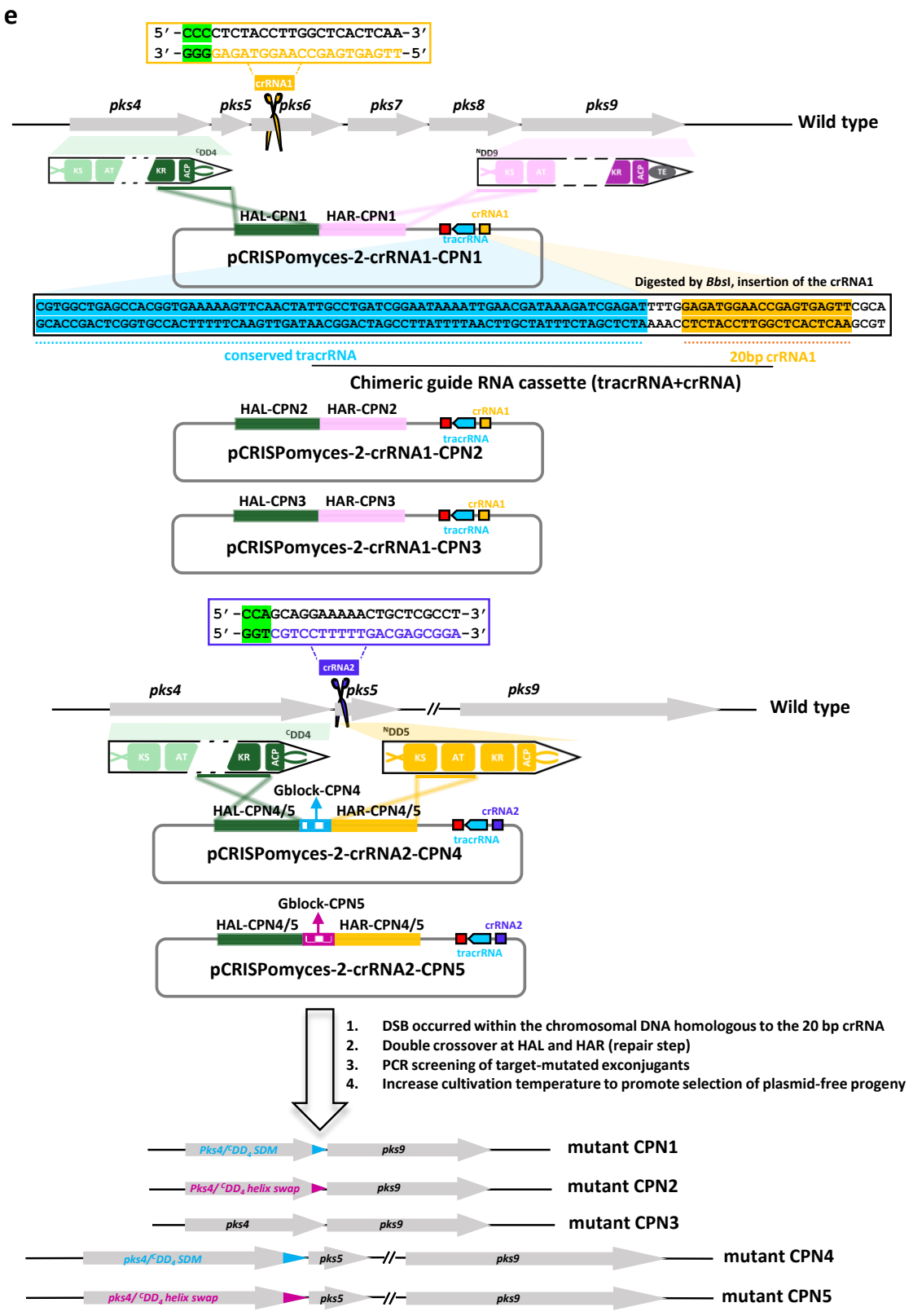
Supplementary Fig. 4 Illustration of the strategy used to generate docking domain engineering constructs K7N1–6 and CPN1–5. **a** Scheme showing the organization of the cosmids used in this study. **b** Plasmids pIJ773, pIJ778, pIJ776 and pSPM88T were used as template for amplifying the cassettes "*FRT+resistance gene+oriT+FRT*" and "*aatL+ resistance gene+oriT+aatR*" used for PCR-targeting. pOSK1111 and pUWL-*oriT-flp* were the relevant plasmids used for cassette excision of "*aatL+resistance gene+oriT+aatR*" and "*FRT+resistance gene+oriT+FRT*", respectively. **c** Delivery plasmids pCRISPomyces-2[4] and pCRISPR-Cas9[5] were employed to conduct CRISPR-Cas9-mediated genetic manipulation.

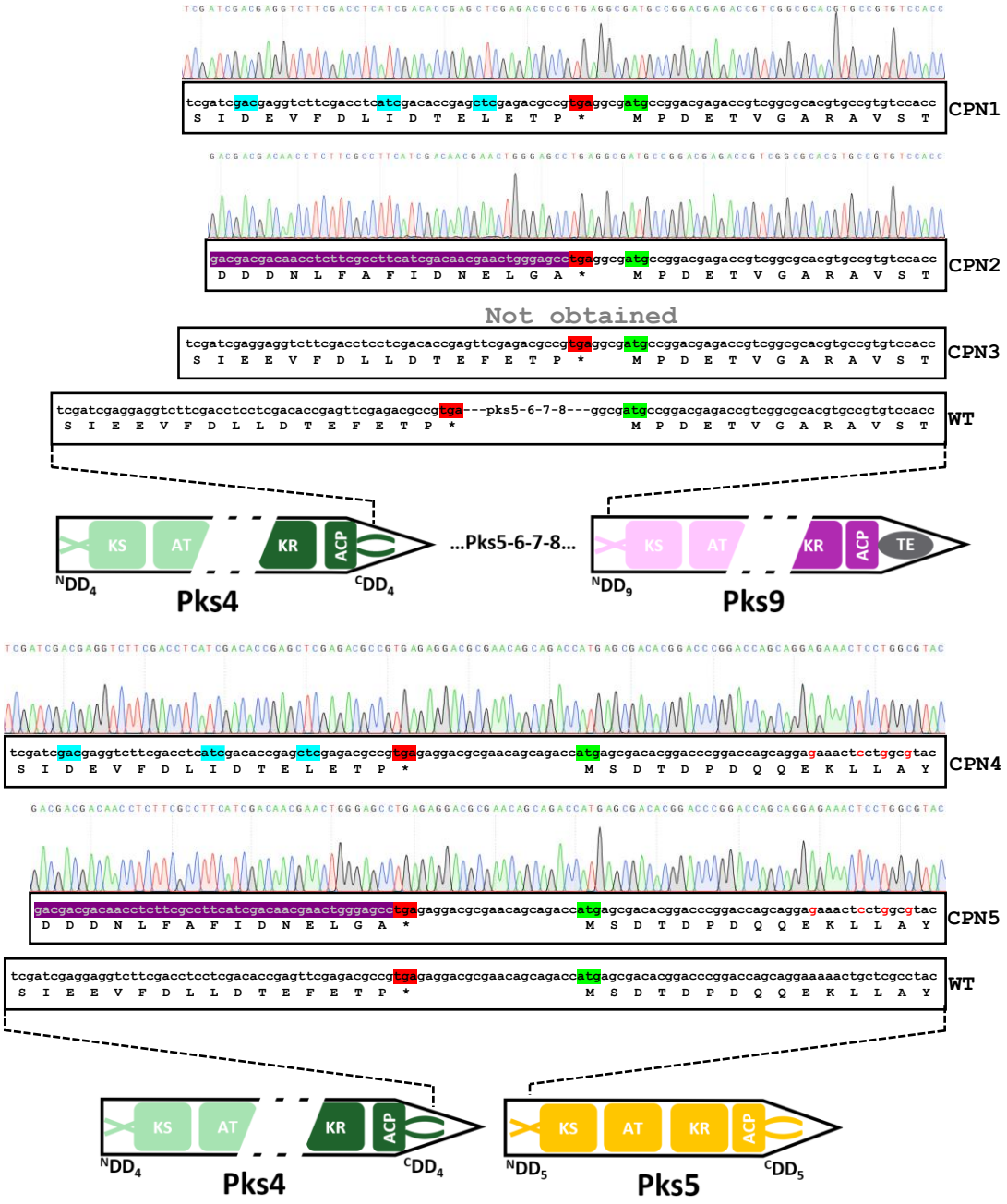
d


BAA9ZA8
+
neo
PCR fragment (PCR-neo)
PCR amplification
110
neo
111
Transformation of E. coli BW25113/pKD20
(λ RED recombination plasmid)
with BAA9ZA8 and PCR-neo, separately
E. coli BW25113/pKD20/BAA9ZA8
target genes
BAA9ZA8
cat
neo
pKD20
Selection of ampicillin and kanamycin resistant transformants
E. coli BW25113/pKD20/BAC1
target genes
BAC1
neo
pKD20
9996
aatL
aac(3)IV+oriT
aatR
9997
PCR amplification
Successive transformation
PCR fragment (PCR-K7N1)
neo
BAC1
target genes
pKD20
aatL
aac(3)IV+oriT
aatR
Incubation overnight at 37°C to promote the loss of pKD20,
selection of apramycin and kanamycin resistant clones.
Extraction of BAC1_K7N1
neo
BAC1_K7N1
aatL
aac(3)IV+oriT
aatR
Transformation of E. coli ET12567/pUZ8002
E. coli ET12567/pUZ8002/BAC1_K7N1
Conjugation with S. ambofaciens ATCC
neo
BAC1_K7N1
aatL
aac(3)IV+oriT
aatR
target genes
ATCC (wild type)
Double crossover, selection of apramycin resistant exconjugants
aatL
aac(3)IV+oriT
aatR
Mutant ATCC/K7N1_aac(3)IV+oriT
Conjugation with E. coli ET12567/pUZ8002/pOSK1111
aatL
aac(3)IV+oriT
aatR
pSAM2 (excisionase and integrase)-induced excision,
selection of apramycin sensitive clones
33bp
Mutant ATCC/K7N1
TCGCGCGCGCTTCGTTCGGGACGAAGAGGCAGA
Conjugation with E. coli ET12567/pUZ8002/pOE484
33bp
ermEp*+samR0484
Mutant K7N1/OE484



Supplementary Fig. 4 Illustration of the strategy used to generate docking domain engineering constructs K7N1–6 and CPN1–5, cont. **d** Schematic representation of construction of K7N1–N6 (represented by K7N1) by PCR-targeting, as well as verification of all constructs by sequencing. The sequences colored in turquoise and purple correspond, respectively, to the SDM (site-directed mutagenesis) and helix-swapped regions of the Pks4 C-terminal docking domain. The locations of the start (green highlighted 'atg') and stop (red highlighted 'tga') codons are also indicated. The residual '33 bp scar' left following PCR-targeting is highlighted in yellow.


e
5′-CCCCTCTACCTTGGCTCACTCAA-3′
3′-GGGGAGATGGAACCGAGTGAGTT-5′
crRNA1
pks4
pks5
pks6
pks7
pks8
pks9
Wild type
CDD4
NDD9
KS
AT
KR
ACP
TE
HAL-CPN1
HAR-CPN1
crRNA1
tracrRNA
pCRISPomyces-2-crRNA1-CPN1
Digested by BbsI, insertion of the crRNA1
CGTGGCTGAGCCACGGTGAAAAAGTTCAACTATTGCCTGATCGGAATAAAATTGAACGATAAAGATCGAGATTTTGGAGATGGAACCGAGTGAGTTCGCA
GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCTCTACCTTGGCTCACTCAAGCGT
conserved tracrRNA
20bp crRNA1
Chimeric guide RNA cassette (tracrRNA+crRNA)
HAL-CPN2
HAR-CPN2
pCRISPomyces-2-crRNA1-CPN2
HAL-CPN3
HAR-CPN3
pCRISPomyces-2-crRNA1-CPN3
5′-CCAGCAGGAAAAACTGCTCGCCT-3′
5′-GGTCGTCCTTTTTGACGAGCGGA-3′
crRNA2
pks4
pks5
pks9
Wild type
CDD4
NDD5
Gblock-CPN4
HAL-CPN4/5
HAR-CPN4/5
crRNA2
tracrRNA
pCRISPomyces-2-crRNA2-CPN4
Gblock-CPN5
HAL-CPN4/5
HAR-CPN4/5
crRNA2
tracrRNA
pCRISPomyces-2-crRNA2-CPN5
1. DSB occurred within the chromosomal DNA homologous to the 20 bp crRNA
2. Double crossover at HAL and HAR (repair step)
3. PCR screening of target-mutated exconjugants
4. Increase cultivation temperature to promote selection of plasmid-free progeny
Pks4/CDD4 SDM
pks9
mutant CPN1
Pks4/CDD4 helix swap
pks9
mutant CPN2
pks4
pks9
mutant CPN3
pks4/CDD4 SDM
pks5
pks9
mutant CPN4
pks4/CDD4 helix swap
pks5
pks9
mutant CPN5



Supplementary Fig. 4 Illustration of the strategy used to generate docking domain engineering constructs K7N1–6 and CPN1–5, cont. **e** Schematic representation of CRISPR-Cas9-mediated mutant construction of CPN1–CPN5, and verification of all constructs by sequencing. The significance of the turquoise, purple, green and red highlighted sequences is as in panel **d**. The additional regions in red font (in CPN4 and CPN5) indicate the nucleotides mutated within the designed crRNA2 in order to avoid unwanted cleavage by Cas9 within the genome of the newly-obtained mutant.

a

The calculated and observed *m/z* values of all metabolites in this study.

Compound	Chemical formula	Retention time (min)	Calculated *m/z* value $[M+2H]^{2+}$	Observed *m/z* value $[M+2H]^{2+}$	Calculated *m/z* value $[M+H]^{+}$	Observed *m/z* value $[M+H]^{+}$	Calculated *m/z* value $[M+Na]^{+}$	Observed *m/z* value $[M+Na]^{+}$
Stambomycin **1**A/B[a]	$C_{73}H_{133}NO_{22}$	22.44	688.9736	688.9697 (−5.66 ppm)[f]	1376.9392	1376.9358 (−2.47 ppm)	1398.9217	n.d.
Stambomycin **1**C/D[a]	$C_{72}H_{131}NO_{22}$	21.44	681.9659	681.9620 (−5.72 ppm)	1362.9236	1362.9207 (−2.13 ppm)	1384.9061	n.d.
Stambomycin **1**E[a]	$C_{71}H_{129}NO_{22}$	21.02	674.9539	674.9531 (−1.19 ppm)	1348.9079	1348.9058 (−1.56 ppm)	1370.8904	n.d.
Stambomycin **1**F[a]	$C_{74}H_{135}NO_{22}$	23.91	694.9815	n.d.	1390.9549	1390.9515 (−2.44 ppm)	1412.9373	n.d.
Shunt (M12) **2**[b]	$C_{35}H_{68}O_{11}$	21.07	333.2452	n.d.	665.4834	665.4831 (−0.45 ppm)	687.4650	687.4649 (−0.15 ppm)
Shunt (M12) **3**[b]	$C_{36}H_{70}O_{11}$	22.25	340.2533	n.d.	679.4991	679.4986 (−0.74 ppm)	701.4812	701.4805 (−1.0 ppm)
Shunt (M13) **4**[b]	$C_{38}H_{72}O_{11}$	23.38	353.2607	n.d.	705.5147	705.5144 (−0.43 ppm)	727.4972	727.4961 (−1.51 ppm)
Shunt (M13) **5**[b]	$C_{39}H_{74}O_{11}$	24.37	360.1495	n.d.	719.5304	719.5299 (−0.69 ppm)	741.5129	741.5115 (−1.89 ppm)
Dehydroxylated shunt (M12) **6**[c]	$C_{35}H_{68}O_{10}$	24.88	325.2485	n.d.	649.4889	649.4858 (−4.8 ppm)	661.4880	n.d.
Dehydroxylated shunt (M12) **7**[c]	$C_{36}H_{70}O_{10}$	26.11	332.2565	n.d.	663.5051	663.5018 (−5.0 ppm)	675.5042	n.d.
Dehydroxylated shunt (M13) **8**[c]	$C_{38}H_{72}O_{10}$	27.18	345.2639	n.d.	689.5198	689.5177 (−3.0 ppm)	711.5023	711.4996 (−3.79 ppm)
Dehydroxylated shunt (M13) **9**[c]	$C_{39}H_{74}O_{10}$	28.27	352.2717	n.d.	703.5355	703.5334 (−3.0 ppm)	725.5180	725.5155 (−3.45 ppm)
Linear ministambomycin **10**A/B[d]	$C_{49}H_{92}O_{15}$	25.38	461.3294	n.d.	921.6509	921.6499 (−1.09 ppm)	943.6334	n.d.
Linear ministambomycin **10**C/D[d]	$C_{48}H_{90}O_{15}$	24.45	454.3216	n.d.	907.6352	907.6346 (−0.66 ppm)	929.6177	n.d.
Non-hydroxylated cyclic ministambomycin **11**A/B[d]	$C_{49}H_{90}O_{14}$	25.37	452.3241	n.d.	903.6403	903.6388 (−1.66 ppm)	925.4428	n.d.
Non-hydroxylated cyclic ministambomycin **11**C/D[d]	$C_{48}H_{88}O_{14}$	24.44	445.3159	n.d.	889.6247	889.6237 (−1.12 ppm)	911.6072	n.d.
Hydroxylated cyclic ministambomycin **12**A/B[d]	$C_{49}H_{90}O_{15}$	25.17	460.3216	n.d.	919.6352	919.6348 (−0.43 ppm)	941.6177	n.d.
Hydroxylated cyclic ministambomycin **12**C/D[d]	$C_{48}H_{88}O_{15}$	24.15	453.3138	n.d.	905.6196	905.6189 (−0.77 ppm)	927.6021	n.d.
Dehydroxylated linear ministambomycin **13**A/B[e]	$C_{49}H_{92}O_{14}$	29.14	453.3320	n.d.	905.6560	905.6543 (−1.88 ppm)	927.6358	n.d.
Dehydroxylated linear ministambomycin **13**C/D[e]	$C_{48}H_{90}O_{14}$	28.09	446.3241	n.d.	891.6403	891.6393 (−1.12 ppm)	913.6228	n.d.
Shunt (M21) **14**A/B[d]	$C_{42}H_{80}O_{12}$	27.01	389.2897	n.d.	777.5723	777.5714 (−1.16 ppm)	799.5547	n.d.
Shunt (M21) **14**C/D[d]	$C_{41}H_{78}O_{12}$	27.98	382.2819	n.d.	763.5566	763.5564 (−0.26 ppm)	785.5391	n.d.
Dehydroxylated shunt (M21) **15**A/B[e]	$C_{42}H_{80}O_{11}$	31.06	381.2923	n.d.	761.5774	761.5765 (−1.17 ppm)	783.5598	n.d.
Dehydroxylated shunt (M21) **15**C/D[e]	$C_{41}H_{78}O_{11}$	29.20	374.2845	n.d.	747.5617	747.5611 (−0.80 ppm)	769.5442	n.d.

n.d. : not detected.

[a]: Retention time and observed masses representive of ATCC/OE484, see **Fig. 3a**.

[b]: Retention time and observed masses representive of CPN2/OE484, see **Figs. 3b and c**. Delayed retention times for these compounds was obsereved during analysis of CPN2/OE484/ACP_{13} SDM (**Supplementary Fig. 9**) due to a change in the HPLC-MS.

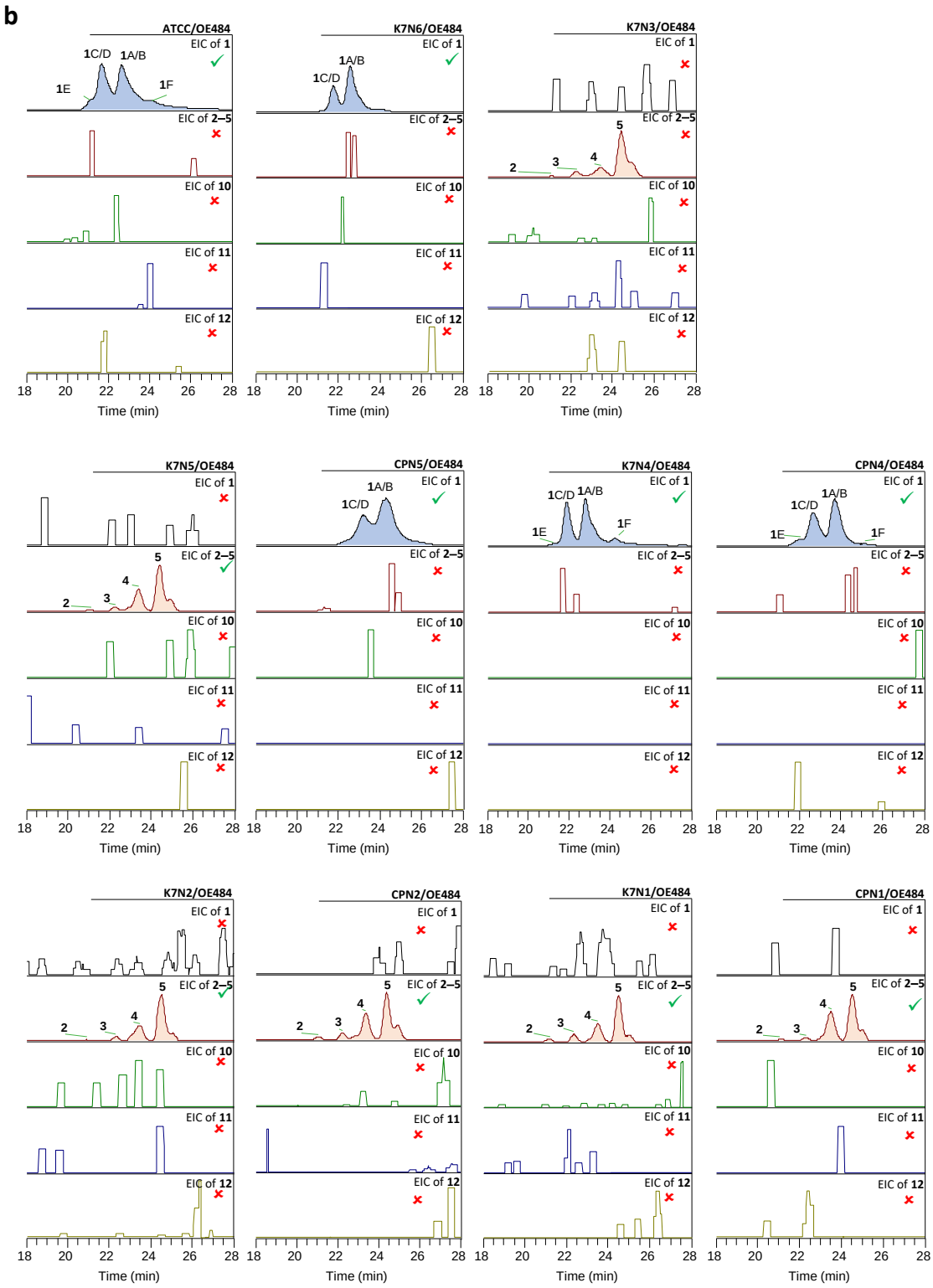
[c]: Retention time and observed masses representive of CPN2/OE484/Δ479, see **Fig. 3c**.

[d]: Retention time and observed masses representive of ATCC/OE484/hy59_S2, see **Supplementary Fig. 10**.

[e]: Retention time and observed masses representive of ATCC/OE484/hy59_S2/Δ479, see **Supplementary Fig. 10**.

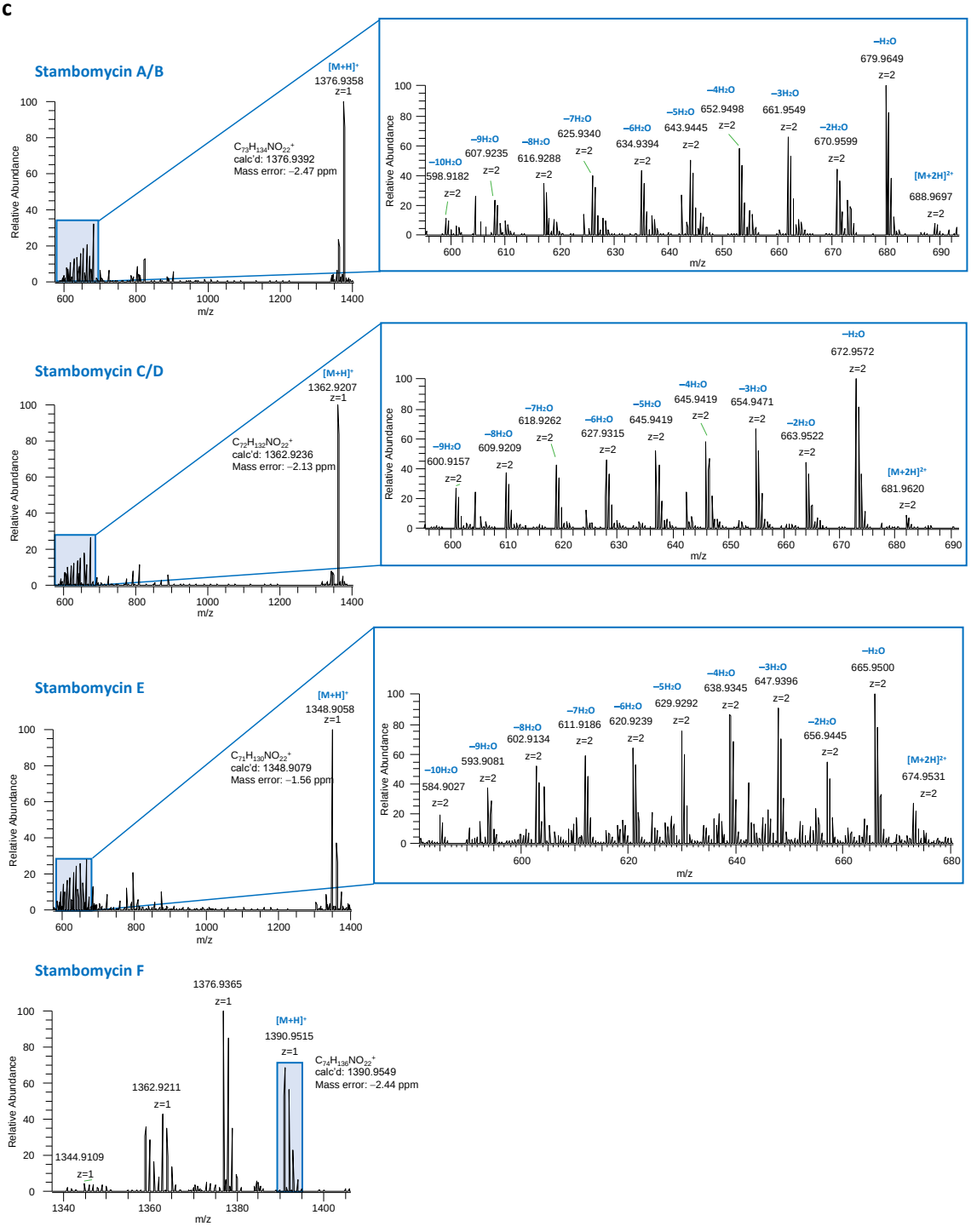
[f]: The values in parentheses indicate the mass errors between the theoretical (calculated) *m/z* and the experimentally observed *m/z*.

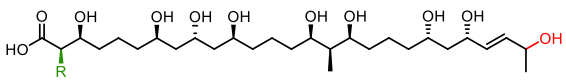
Supplementary Fig. 5 HPLC-MS analysis of all docking domain engineering mutants. a Table summarizing the HPLC-MS retention times, calculated *m/z* values, and high-resolution observed *m/z* values (with errors) for all key metabolites.



Supplementary Fig. 5 HPLC-MS analysis of all docking domain engineering mutants, cont. **b** EIC (extracted ion chromatogram) analysis for the presence of stambomycins **1**, shunt products **2**–**5** and mini-stambomycins **10**–**12**, in extracts of all constructs, by comparison to wild type. The EICs represent all possible *m*/*z* values for each compound family listed in **a**.

c

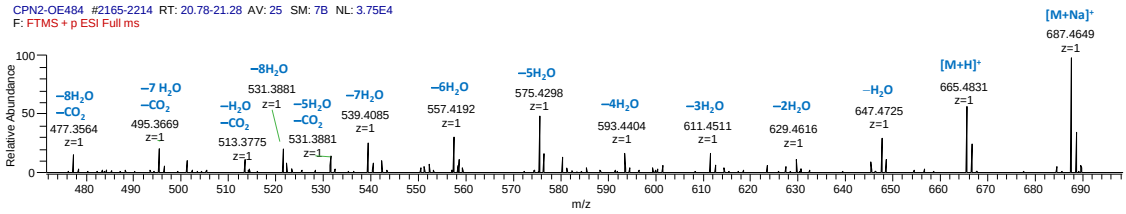

Stambomycin A/B
[M+H]+
1376.9358
z=1
C73H134NO22+
calc'd: 1376.9392
Mass error: −2.47 ppm
−H2O
679.9649
z=2
−4H2O
652.9498
z=2
−3H2O
661.9549
z=2
−5H2O
643.9445
z=2
−7H2O
625.9340
z=2
−6H2O
634.9394
z=2
−2H2O
670.9599
z=2
−9H2O
607.9235
z=2
−8H2O
616.9288
z=2
−10H2O
598.9182
z=2
[M+2H]2+
688.9697
z=2
Relative Abundance
m/z
Stambomycin C/D
[M+H]+
1362.9207
z=1
C72H132NO22+
calc'd: 1362.9236
Mass error: −2.13 ppm
−H2O
672.9572
z=2
−4H2O
645.9419
z=2
−3H2O
654.9471
z=2
−5H2O
645.9419
z=2
−7H2O
618.9262
z=2
−6H2O
627.9315
z=2
−2H2O
663.9522
z=2
−8H2O
609.9209
z=2
−9H2O
600.9157
z=2
[M+2H]2+
681.9620
z=2
Relative Abundance
m/z
Stambomycin E
[M+H]+
1348.9058
z=1
C71H130NO22+
calc'd: 1348.9079
Mass error: −1.56 ppm
−H2O
665.9500
z=2
−4H2O
638.9345
z=2
−3H2O
647.9396
z=2
−5H2O
629.9292
z=2
−7H2O
611.9186
z=2
−6H2O
620.9239
z=2
−2H2O
656.9445
z=2
−8H2O
602.9134
z=2
−9H2O
593.9081
z=2
−10H2O
584.9027
z=2
[M+2H]2+
674.9531
z=2
Relative Abundance
m/z
Stambomycin F
1376.9365
z=1
[M+H]+
1390.9515
z=1
C74H136NO22+
calc'd: 1390.9549
Mass error: −2.44 ppm
1362.9211
z=1
1344.9109
z=1
Relative Abundance
m/z



Shunt product (M12) **2**

Chemical Formula: $C_{35}H_{68}O_{11}$ R = CH_2-CH_2-CH_2-CH(CH_3)-CH_3
CH_2-CH_2-CH_2-CH_2-CH_2-CH_3

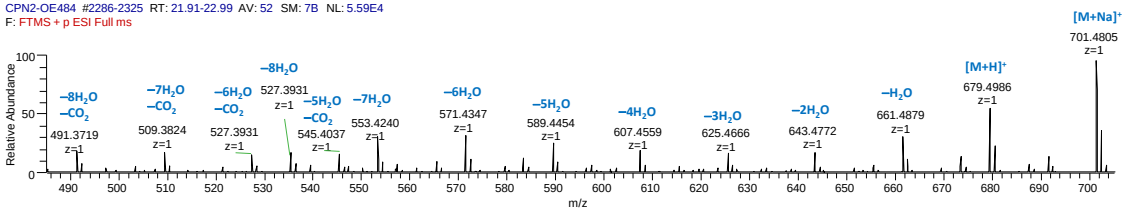
$[M+H]^+$: $C_{35}H_{69}O_{11}^+$, Calc'd: 665.4834, Mass error = –0.45 ppm

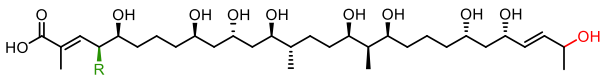


Shunt product (M12) **3**

Chemical Formula: $C_{36}H_{70}O_{11}$ R = CH_2-CH_2-CH_2-CH(CH_3) -CH_2-CH_3
CH_2-CH_2-CH_2-CH_2-CH (CH_3) -CH_3

$[M+H]^+$: $C_{36}H_{71}O_{11}^+$, Calc'd: 679.4991, Mass error = –0.74 ppm

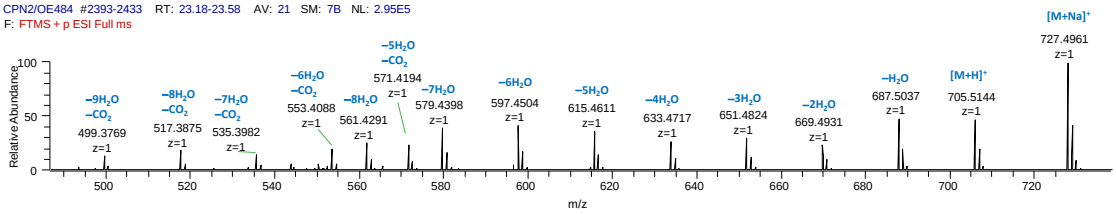




Shunt product (M13) **4**

Chemical Formula: $C_{38}H_{72}O_{11}$ R = CH_2-CH_2-CH_2-CH(CH_3)-CH_3
CH_2-CH_2-CH_2-CH_2-CH_2-CH_3

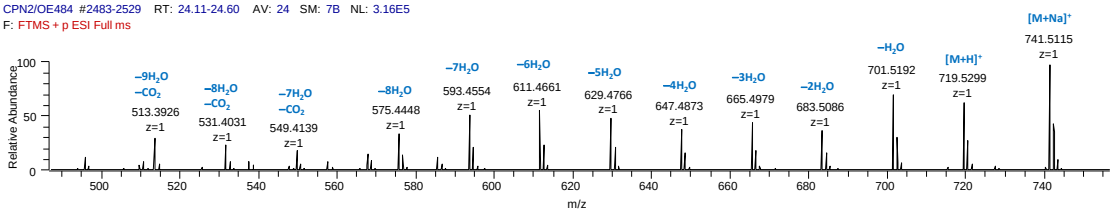
M+H]$^+$: $C_{38}H_{73}O_{11}^+$, Calc'd: 705.5147, Mass error = –0.43 ppm



Shunt product (M13) **5**

Chemical Formula: $C_{39}H_{74}O_{11}$ R = CH_2-CH_2-CH_2-CH(CH_3) -CH_2-CH_3
CH_2-CH_2-CH_2-CH_2-CH (CH_3) -CH_3

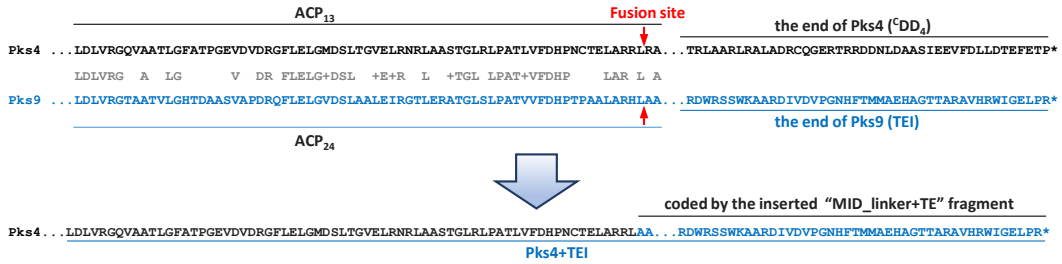
$[M+H]^+$: $C_{39}H_{75}O_{11}^+$, Calc'd: 719.5304, Mass error = –0.69 ppm



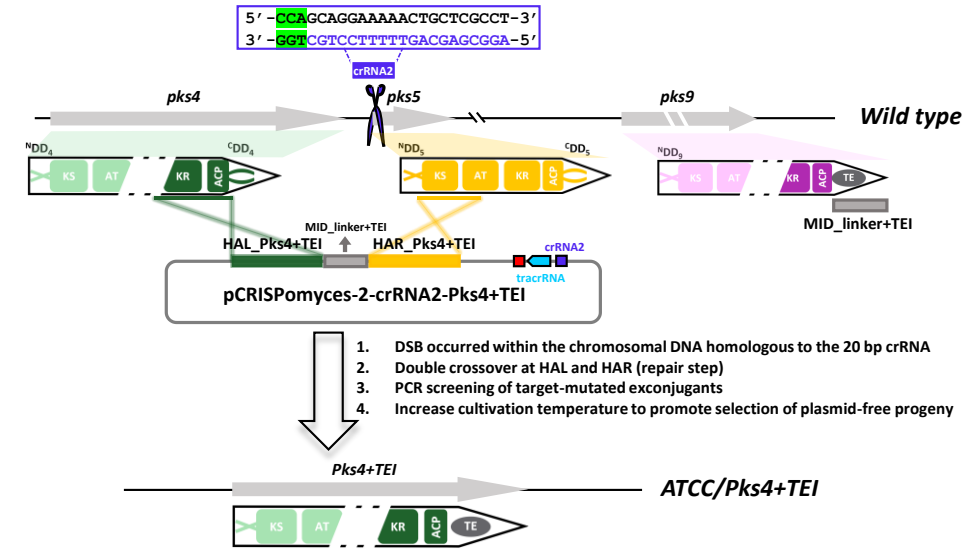
Supplementary Fig. 5 HPLC-MS analysis of all docking domain engineering mutants, cont. **c** Mass spectra of stambomycins **1** (represented by data from analysis of ATCC/OE484) and shunt products **2**–**5** (represented by data from analysis of CPN2/OE484).

a

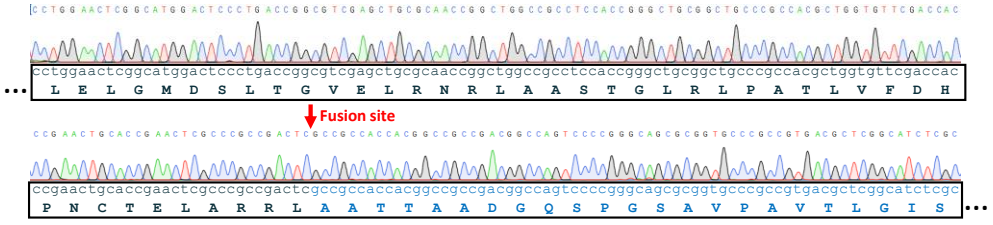
Fusion site for the Pks9 TEI to Pks4:



Cloning strategy:

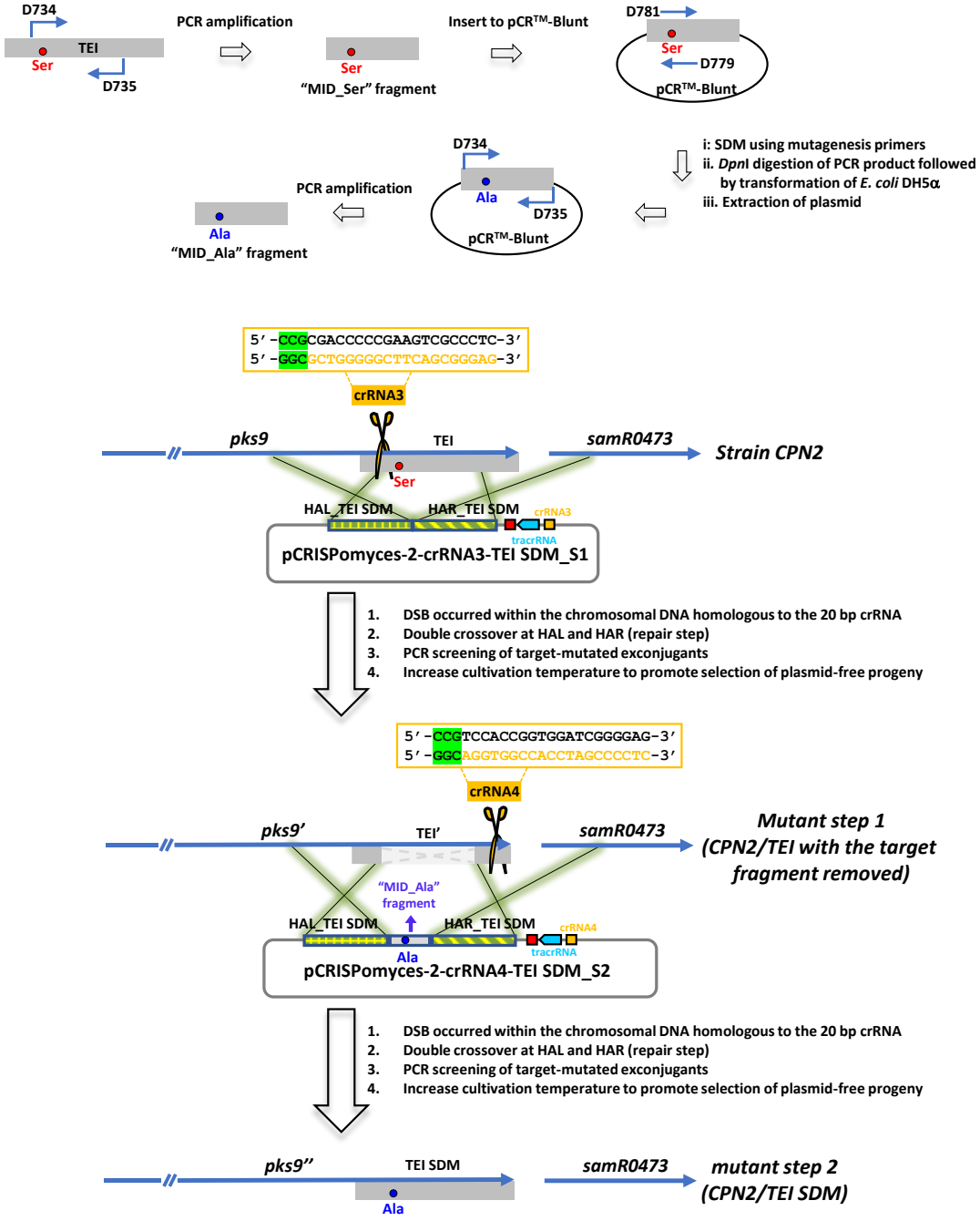


Representative sequencing result of the mutant ATCC/Pks4+TEI:

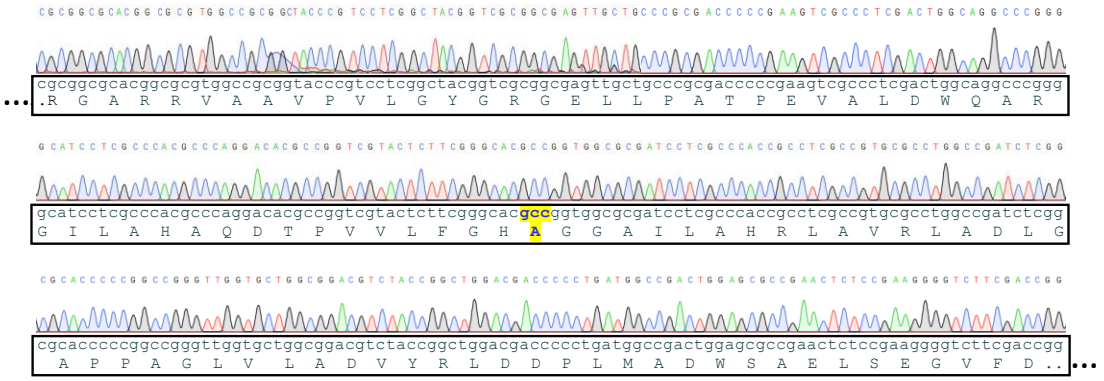


Supplementary Fig. 6 Illustration of the strategy for TE inactivation and TEI fusion. **a** Schematic representation of the covalent fusion of the Pks9 TEI domain to Pks4 (the selected fusion site is highlighted by a red arrow). The sequence of Pks4 is in black and that of Pks9 in blue.

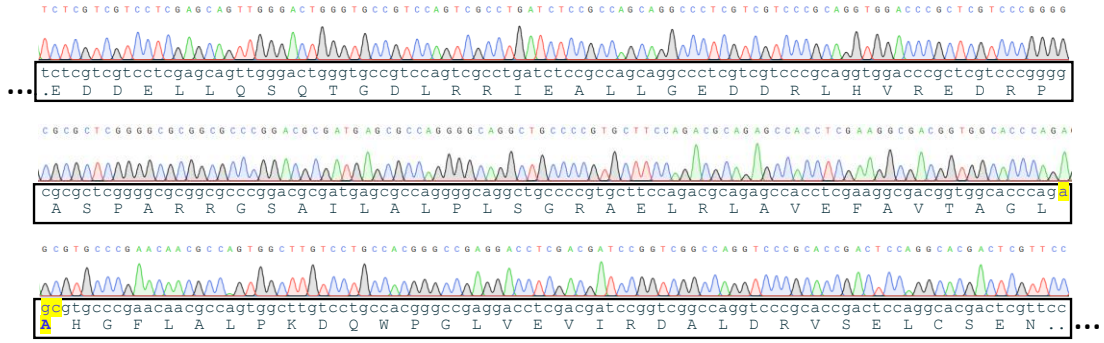
b


D734
TEI
Ser
D735
PCR amplification
Ser
"MID_Ser" fragment
Insert to pCR™-Blunt
D781
Ser
D779
pCR™-Blunt
i: SDM using mutagenesis primers
ii. DpnI digestion of PCR product followed by transformation of E. coli DH5α
iii. Extraction of plasmid
D734
Ala
D735
pCR™-Blunt
PCR amplification
Ala
"MID_Ala" fragment
5′-CCGCGACCCCCGAAGTCGCCCTC-3′
5′-GGCGCTGGGGGCTTCAGCGGGAG-3′
crRNA3
pks9
TEI
samR0473
Strain CPN2
Ser
HAL_TEI SDM
HAR_TEI SDM
crRNA3
tracrRNA
pCRISPomyces-2-crRNA3-TEI SDM_S1
1. DSB occurred within the chromosomal DNA homologous to the 20 bp crRNA
2. Double crossover at HAL and HAR (repair step)
3. PCR screening of target-mutated exconjugants
4. Increase cultivation temperature to promote selection of plasmid-free progeny
5′-CCGTCCACCGGTGGATCGGGGAG-3′
5′-GGCAGGTGGCCACCTAGCCCCTC-3′
crRNA4
pks9′
TEI′
samR0473
Mutant step 1
(CPN2/TEI with the target fragment removed)
"MID_Ala" fragment
HAL_TEI SDM
HAR_TEI SDM
crRNA4
tracrRNA
Ala
pCRISPomyces-2-crRNA4-TEI SDM_S2
1. DSB occurred within the chromosomal DNA homologous to the 20 bp crRNA
2. Double crossover at HAL and HAR (repair step)
3. PCR screening of target-mutated exconjugants
4. Increase cultivation temperature to promote selection of plasmid-free progeny
pks9″
TEI SDM
samR0473
mutant step 2
(CPN2/TEI SDM)
Ala

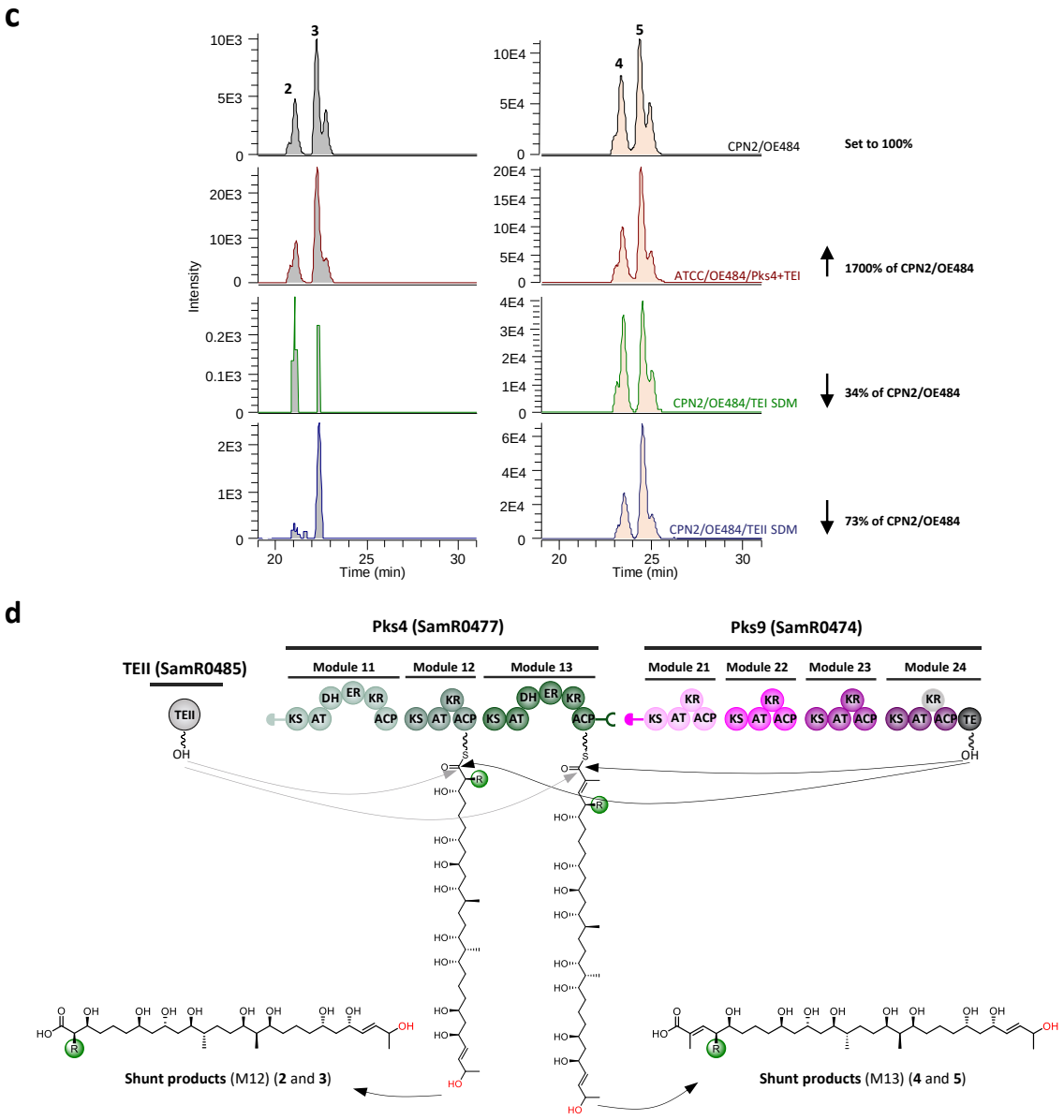
Representative sequencing result of the mutant CPN2/TEI SDM



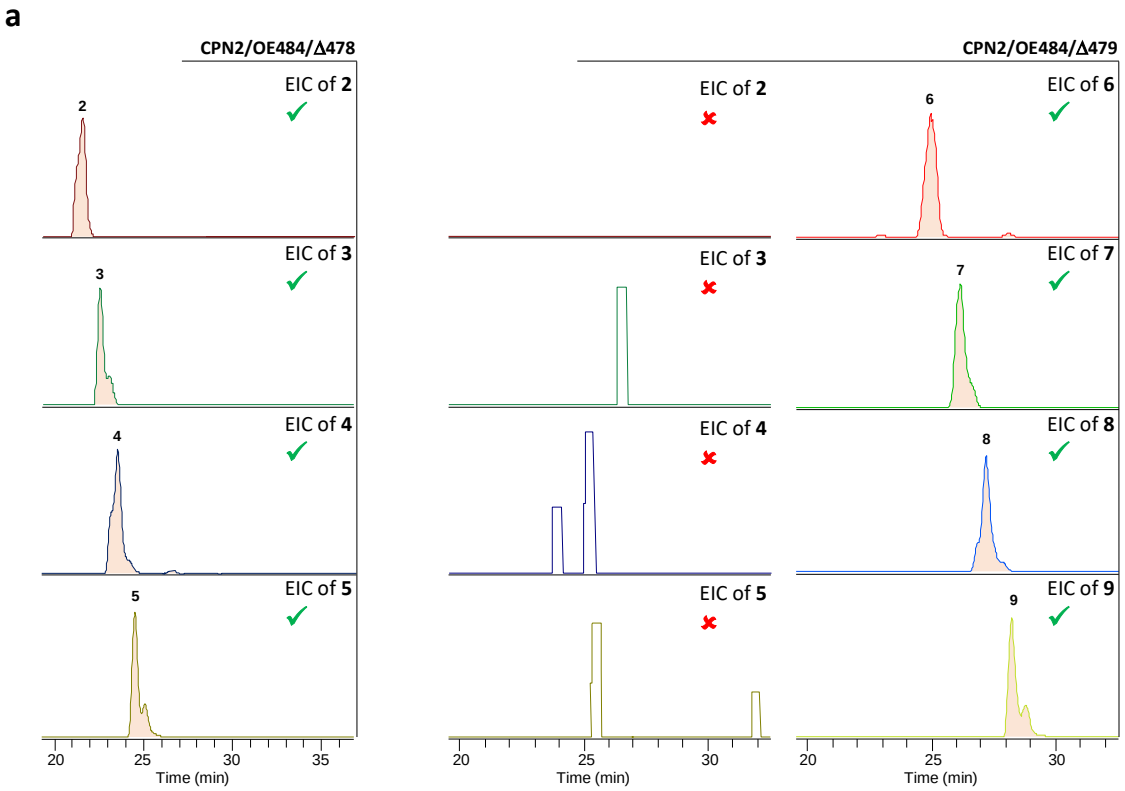
Representative sequencing result of the mutant CPN2/TEII SDM



Supplementary Fig. 6 Illustration of the strategy for TE inactivation and TEI fusion, cont. **b** Schematic representation of TEI (and TEII) domain inactivation in strain CPN2 (represented by TEI inactivation). The catalytic Ser residue was successfully mutated to Ala (mutation verification is highlighted in yellow in the sequencing).

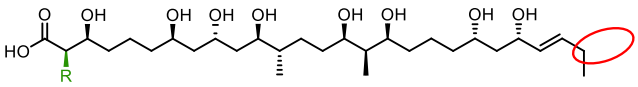


Supplementary Fig. 6 Illustration of the strategy for TE inactivation and TEI fusion, cont. **c** HPLC-MS analysis of mutants ATCC/OE484/Pks4+TEI, CPN2/OE484/TEI SDM and CPN2/OE484/TEII SDM. Based on integrated EIC peak area (see **Supplementary Table 5**), the yields of shunt products (**2**–**5**) in mutant ATCC/Pks4+TEI were increased up to 17-fold relative to CPN2/OE484. Conversely, the inactivation of the TEI domain or TEII reduced the yield of shunt products (**2**–**5**) to 34% and 73%, respectively, relative to that of CPN2/OE484 (set to 100%). **d** Illustration of the model for the off-loading of stambomycin intermediates. A proposed interaction between TEI of Pks9 and Pks4 allows the TEI to directly off-load the immature polyketide from modules 12 and 13 to generate the shunt products (**2**–**5**). The *trans*-acting type II thioesterase (TEII), homologs of which have been shown to exhibit proof-reading activity towards inappropriately decarboxylated extender units and stalled intermediates[6], can also release the shunt products from the ACP domains.



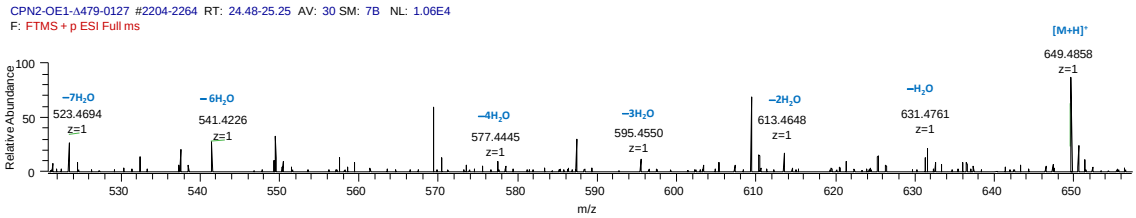
Supplementary Fig. 7 Inactivation of SamR0478 and SamR0479 in strain CPN2. a Inactivation of SamR0478 did not change the metabolic profile compared to mutant CPN2/OE484. Inactivation of SamR0479 in mutant CPN2/OE484 disabled the production of **2**–**5** and led to dehydroxylated shunt products **6–9**.

b



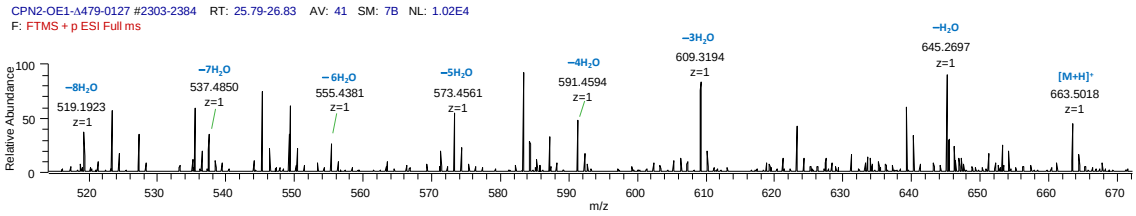
Dehydroxylated shunt (M12) **6**

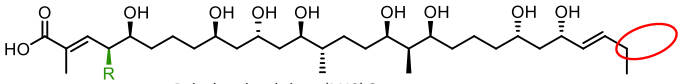
Chemical Formula: $C_{35}H_{68}O_{10}$ R = CH_2-CH_2-CH_2-CH(CH_3)-CH_3
CH_2-CH_2-CH_2-CH_2-CH_2-CH_3
$[M+H]^+$: $C_{35}H_{69}O_{10}^+$, Calc'd: 649.4889, Mass error = –4.8 ppm



Dehydroxylated shunt (M12) **7**

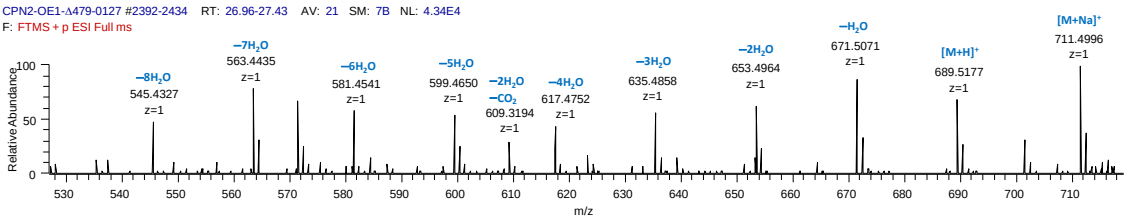
Chemical Formula: $C_{36}H_{70}O_{10}$ R = CH_2-CH_2-CH_2-CH(CH_3) -CH_2-CH_3
CH_2-CH_2-CH_2-CH_2-CH (CH_3) -CH_3
$[M+H]^+$: $C_{36}H_{71}O_{10}^+$, Calc'd: 663.5051, Mass error = –5.0 ppm





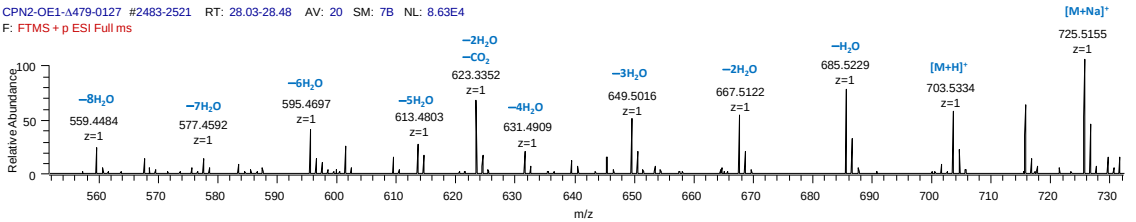
Dehydroxylated shunt (M13) **8**

Chemical Formula: $C_{38}H_{72}O_{10}$ R = CH_2-CH_2-CH_2-CH(CH_3)-CH_3
CH_2-CH_2-CH_2-CH_2-CH_2-CH_3
$[M+H]^+$: $C_{38}H_{73}O_{10}^+$, Calc'd: 689.5198, Mass error = –3.0 ppm



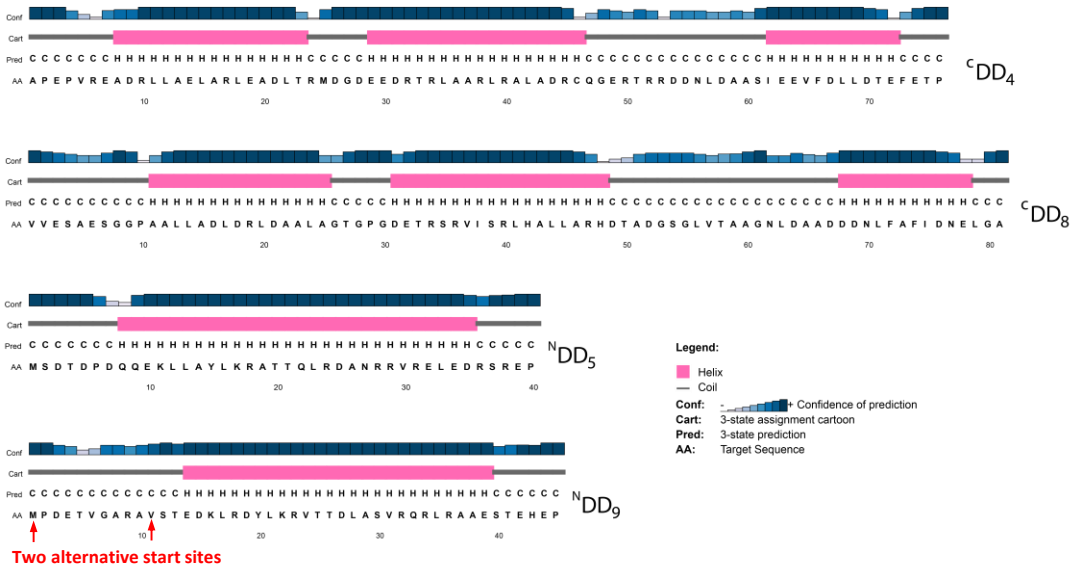
Dehydroxylated shunt (M13) **9**

Chemical Formula: $C_{39}H_{74}O_{10}$ R = CH_2-CH_2-CH_2-CH(CH_3) -CH_2-CH_3
CH_2-CH_2-CH_2-CH_2-CH (CH_3) -CH_3
$[M+H]^+$: $C_{39}H_{75}O_{10}^+$, Calc'd: 703.5355, Mass error = –3.0 ppm

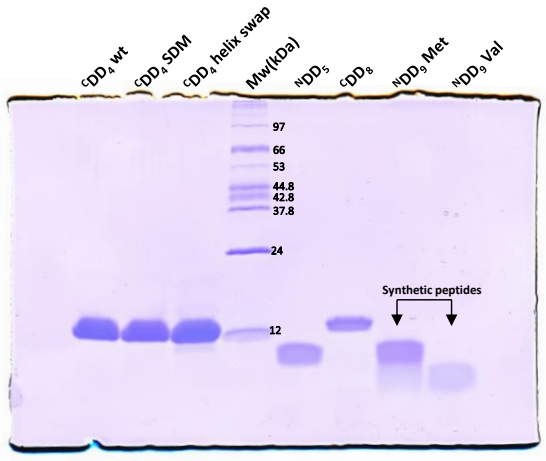


Supplementary Fig. 7 Inactivation of SamR0478 and SamR0479 in strain CPN2, cont. **b** Mass spectra and structures of dehydroxylated shunt products. The red circle on the structure indicates the missing terminal hydroxyl group.

a

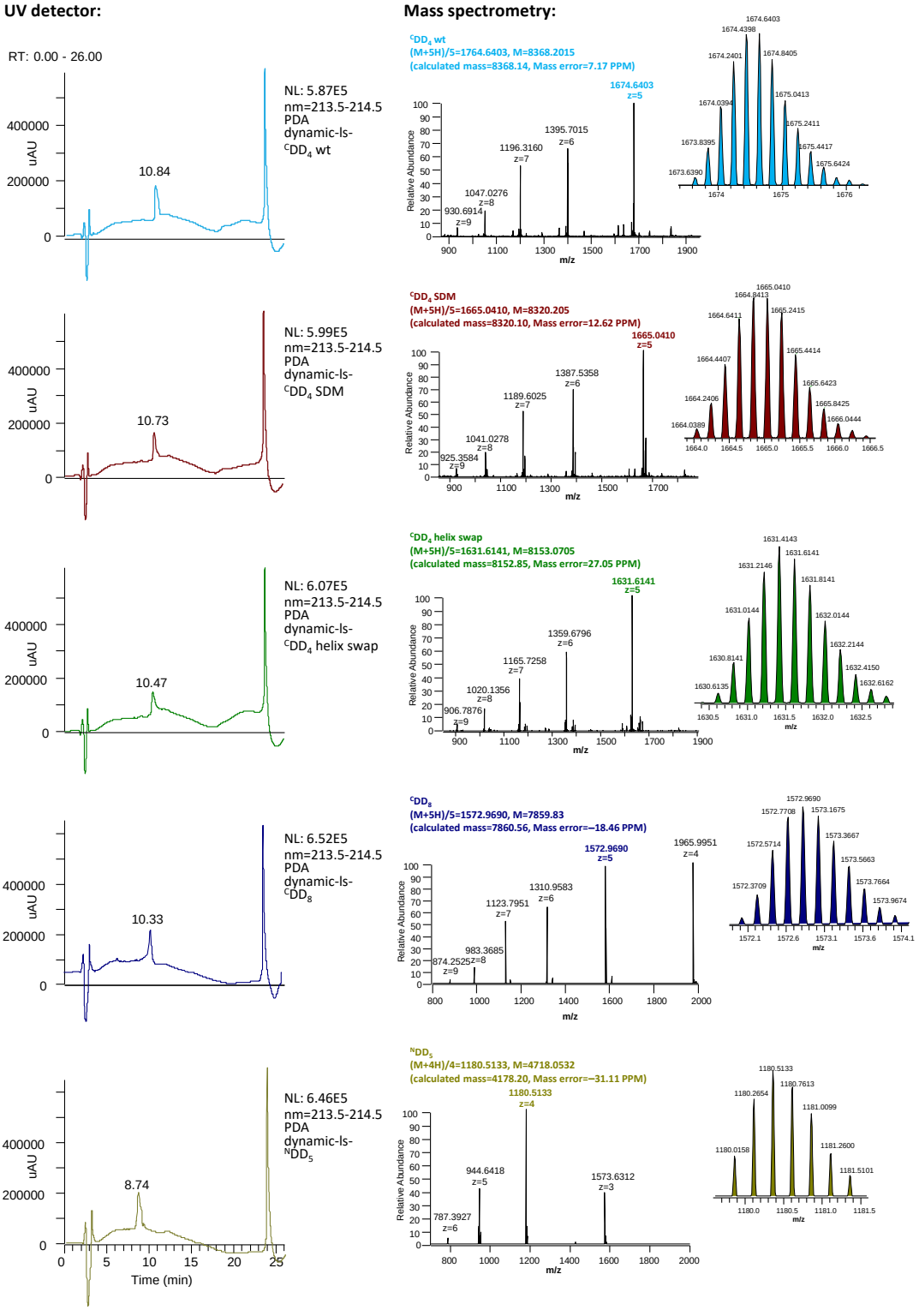


b



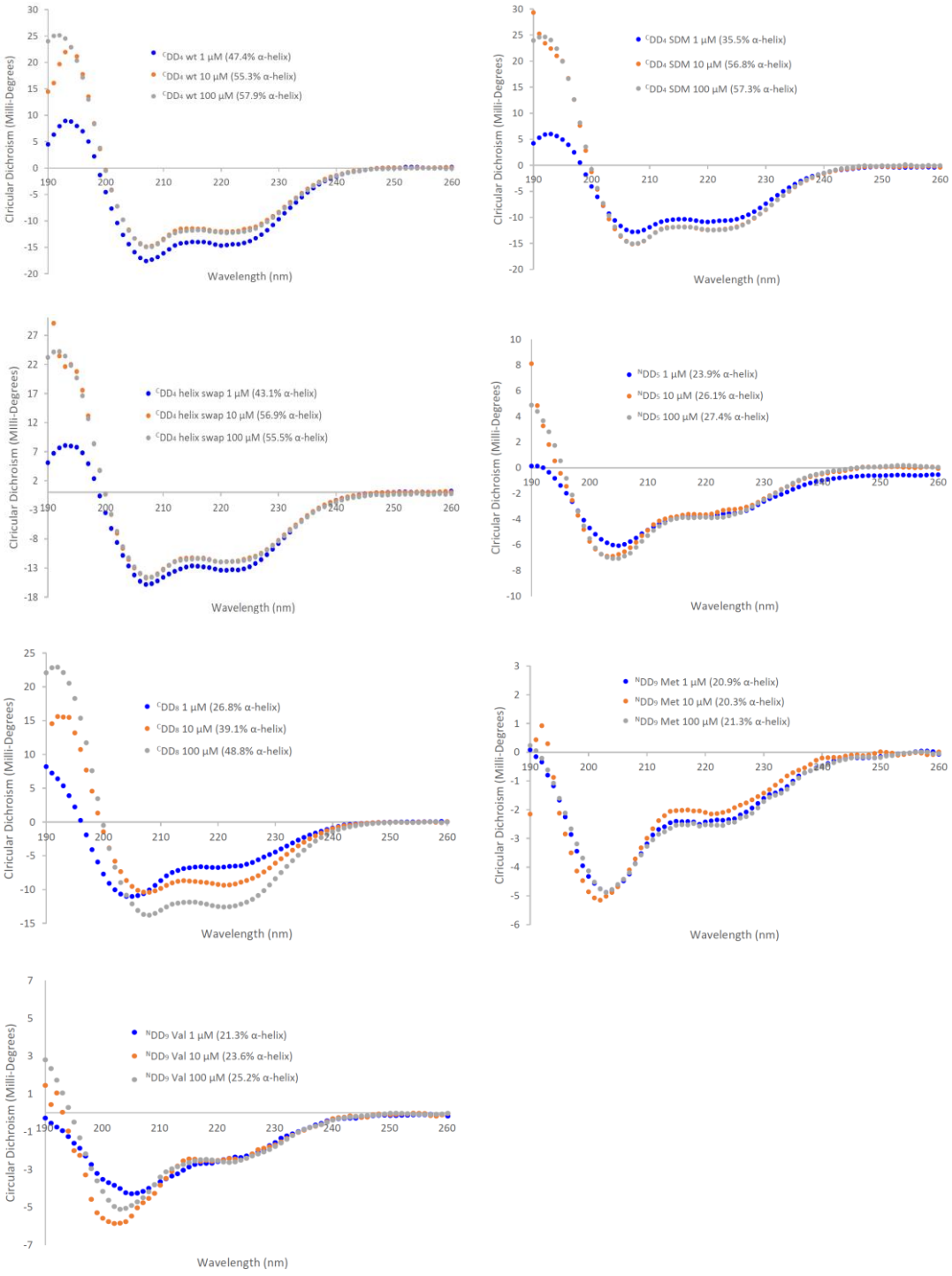
Supplementary Fig. 8 Analysis *in vitro* of docking domains. a PSIPRED graphical output of secondary structure prediction on docking domains. Based on this analysis, we identified an extended random coil region in $^{N}DD_9$ upstream of the predicted coiled-coil structure, which prompted us to speculate that the Val (Val-Ser-Thr…) is likely to be the true start site. **b** SDS-PAGE analysis of docking domains (both recombinant and commercially-synthesized) used in this work.

c



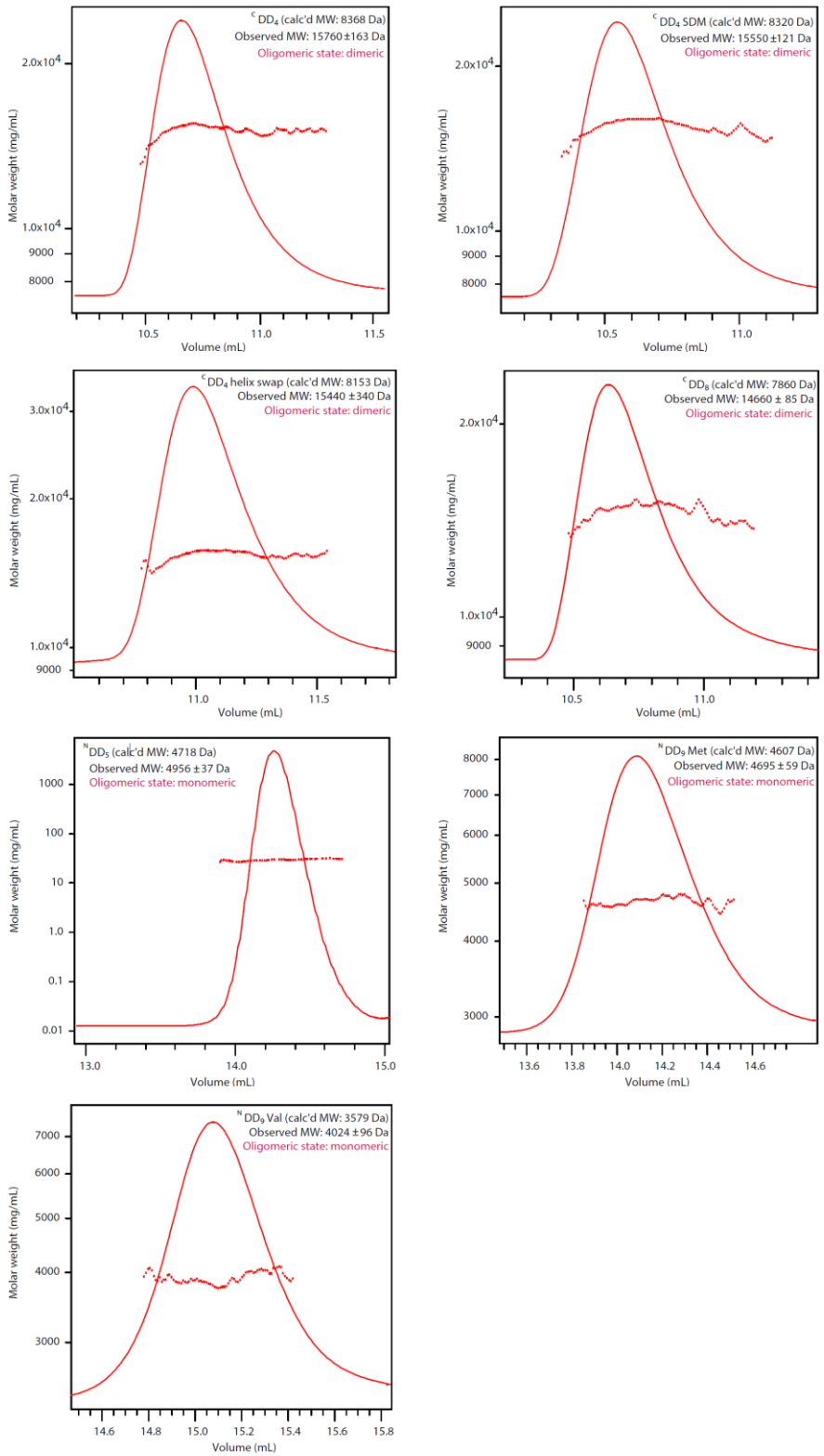
Supplementary Fig. 8 Analysis *in vitro* of docking domains, cont. **c** HPLC and mass spectrometry analysis of recombinant docking domains ($^{C}DD_4$ wt, $^{C}DD_4$ SDM, $^{C}DD_4$ helix swap, $^{N}DD_5$, and $^{C}DD_8$).

d



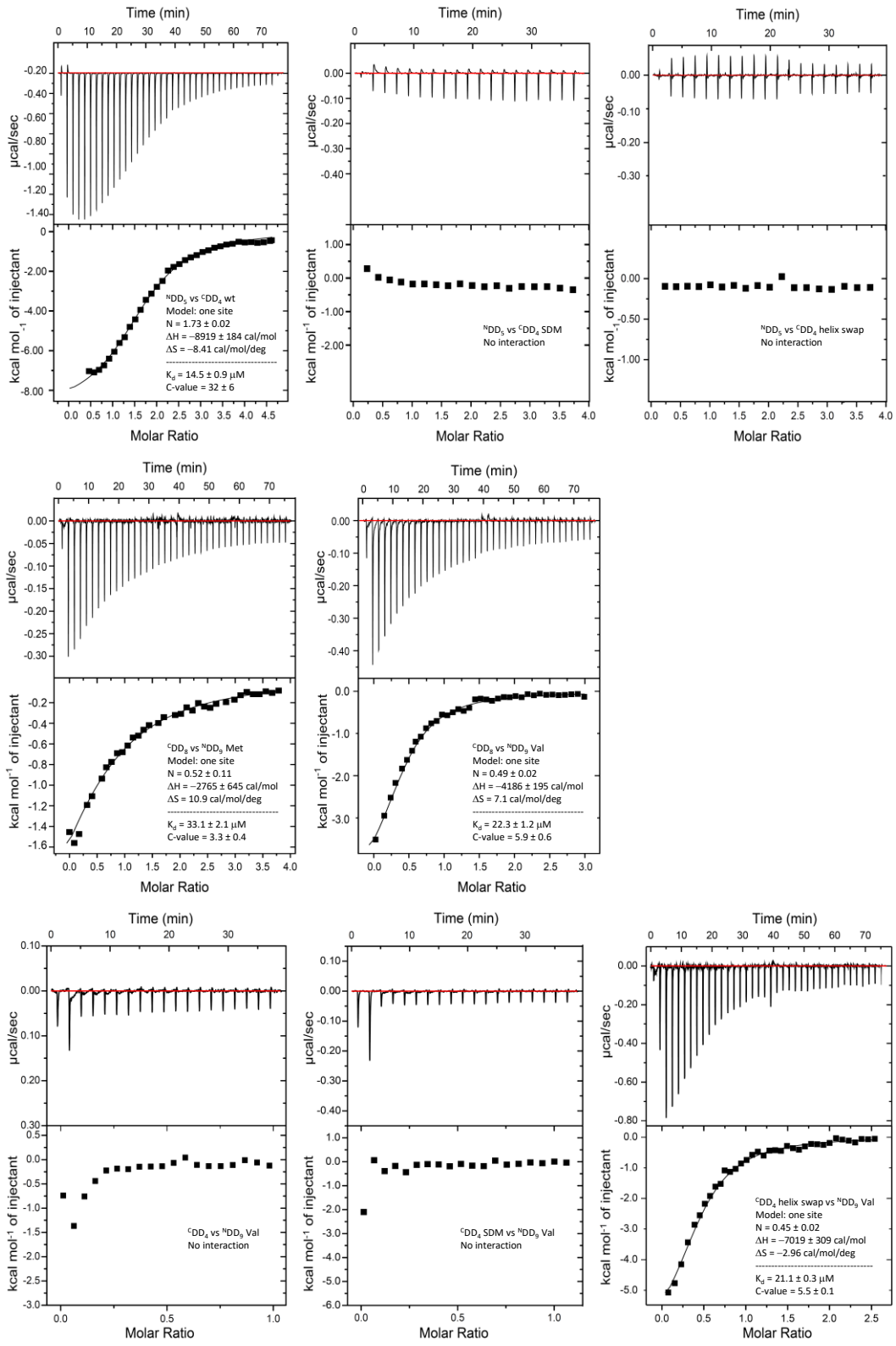
Supplementary Fig. 8 Analysis *in vitro* of docking domains, cont. **d** Circular dichroism analysis of all docking domains at 100, 10 and 1 µM. The percentage α-helices in each case were calculated using deconvolution software CDNN2.1[7].

e



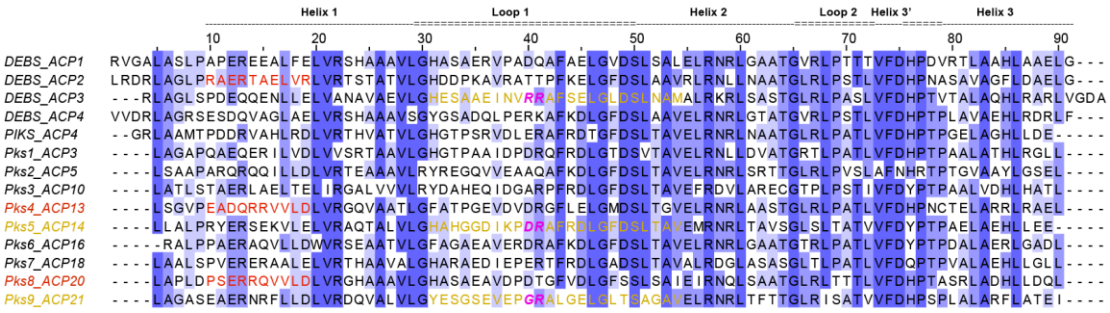
Supplementary Fig. 8 Analysis *in vitro* of docking domains, cont. **e** SEC-MALS analysis of docking domains. Comparison of the observed molecular weights (MW) and the calculated molecular weights (in brackets) shows that all ^{C}DDs are in dimeric in solution and the ^{N}DDs are uniformly monomeric.

f



Supplementary Fig. 8 Analysis *in vitro* of docking domains, cont. **f** ITC thermograms and the calculated binding affinities between pairs of ^{C}DDs and ^{N}DDs. One set of representative data (from two replicate experiments) is shown, but the K_d and C-values represent the average of the two measurements.

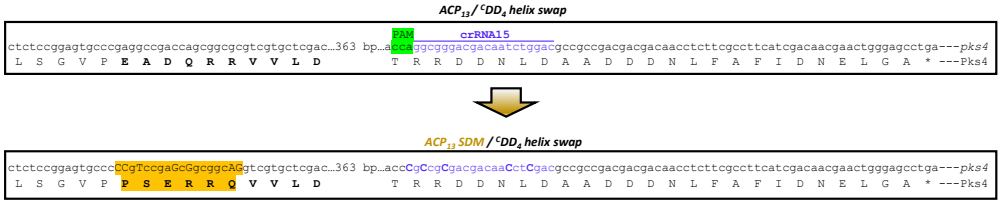
a



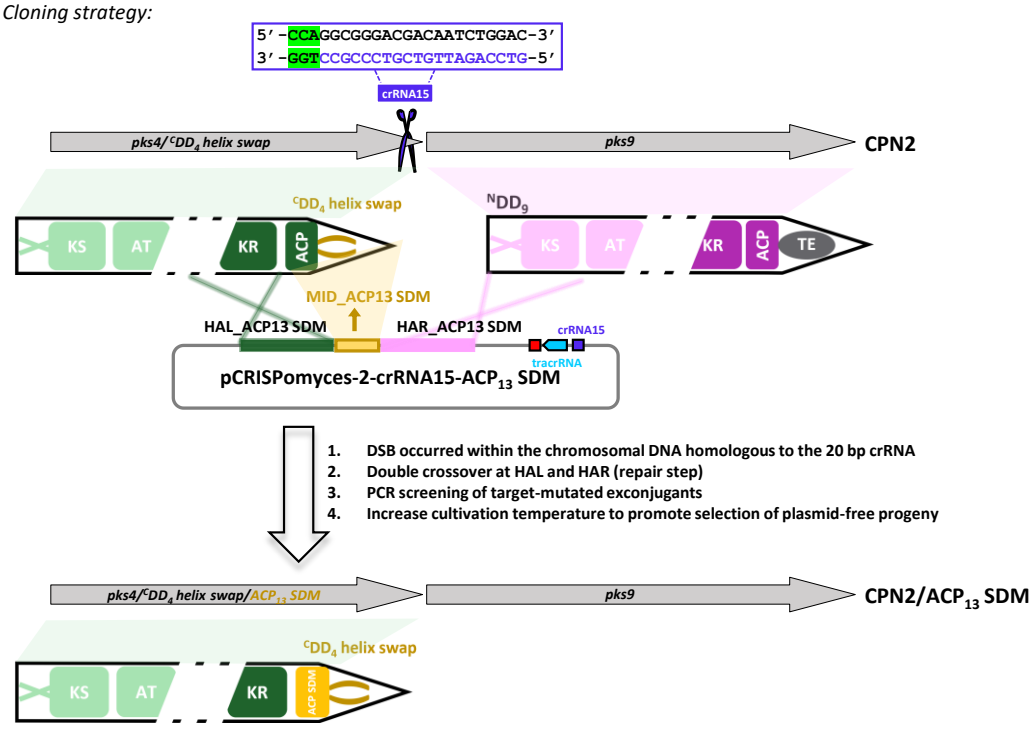
Supplementary Fig. 9 Mutation of the ACP_{13} domain in strain CPN2. **a** Multiple sequence alignment of stambomycin ACP domains to identify the residues/motifs putatively mediating interaction with the down- and upstream KS domains. Assignment of secondary structure and residue numbering is based on the solved NMR structure of DEBS ACP_2 (PDB: 2JU2[8]). Serine 50 is the conserved residue that serves as the attachment point for the phosphopantetheine prosthetic group. Residues 10−19 on helix α1 (H1) (in red) were identified as involved in the ACP/KS intermodular chain translocation[9], while residues 31−54 located on loop 1 and helix α2 (L1+H2 region) (in yellow) were found to contribute to ACP/KS communication during intramodular chain elongation[9]. The bold italic 'RR' on loop 1 of DEBS ACP_3 represents two positions that were identified previously to be most critical to the interaction[10]; the corresponding residues on ACP_{14} and ACP_{21} are 'DR' and 'GR', respectively.

b

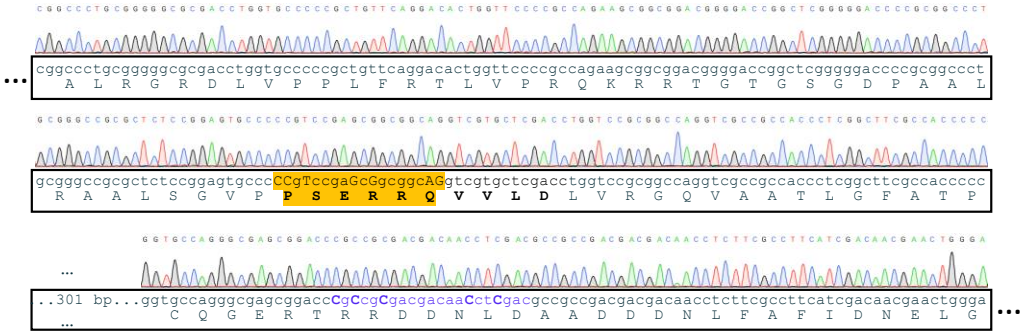
Modification residues of ACP_{13}:



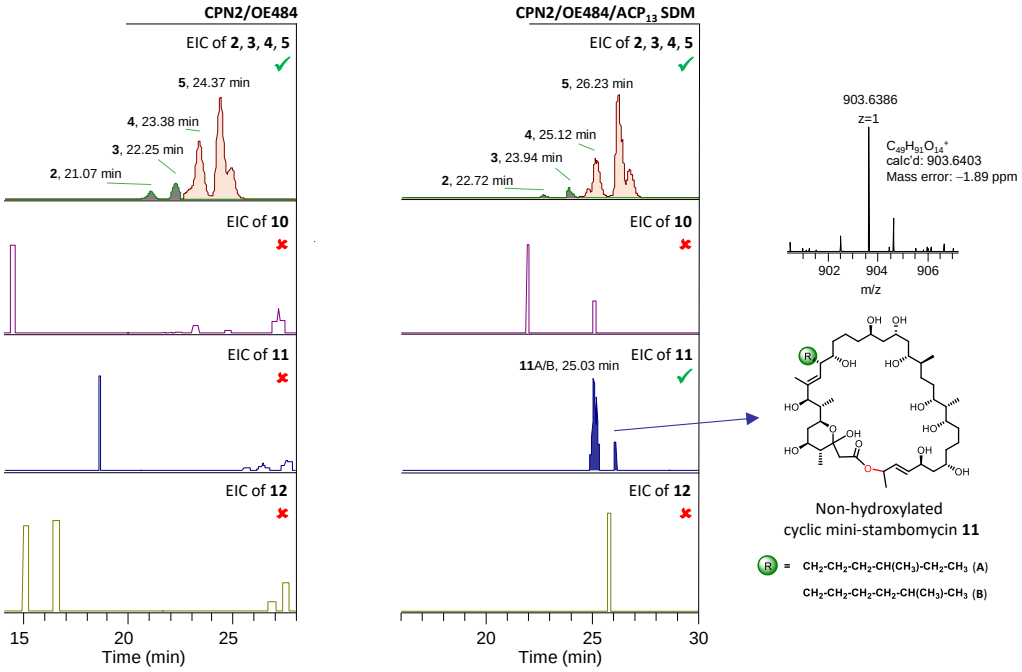
Cloning strategy:



Representative sequencing result of the mutant CPN2/ACP_{13} SDM:

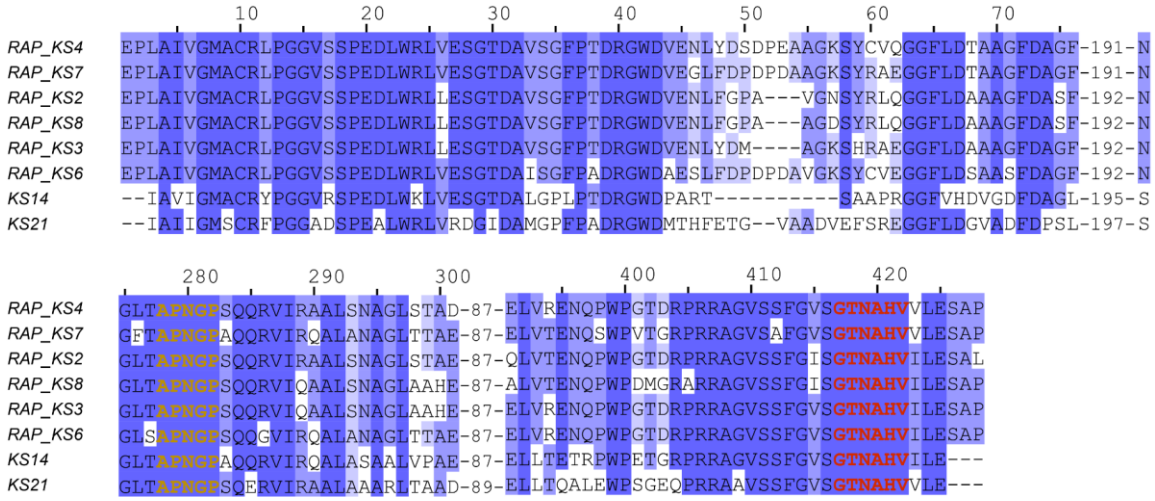


c



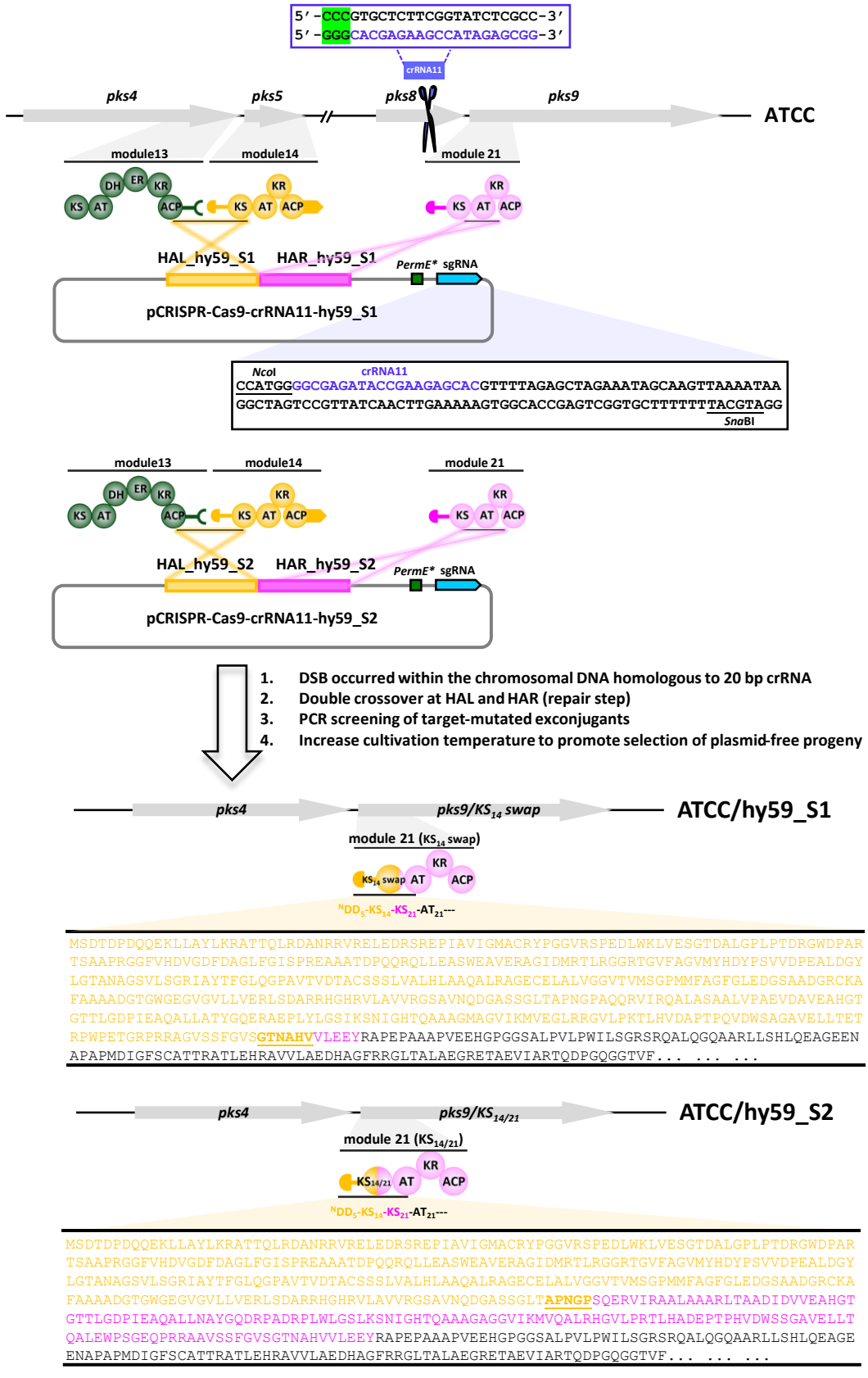
Supplementary Fig. 9 Mutation of the ACP_{13} domain in strain CPN2, cont. **b** Illustration of the strategy to modify ACP_{13} in mutant CPN2. The first 10 residues of helix α1 on ACP_{13} were mutated from "**EADQ**RR**R**VVLD" to "**PSER**R**Q**VVLD" (highlighted in orange). In addition, the DNA sequence corresponding to the selected crRNA15 was also mutated (as shown in blue capital letters) to avoid unwanted cleavage by Cas9 within the genome of the newly-obtained mutant. **c** HPLC-MS analysis of the mutant CPN2/ACP_{13} SDM. An EIC peak with the mass of 903.6386 ($[M+H]^+$), corresponding to the cyclic mini-stambomycins A/B **11** lacking the hydroxyl introduced by SamR0478, was detected successfully. Note: the discrepancy in retention times for shunt metabolites **2–5** between the mutants CPN2/OE484 and CPN2/OE484/ACP_{13} SDM reflects the fact that the analyses were carried out on two different machines, a UHPLC-LTQ Orbitrap and a UHPLC-Orbitrap ID-X Tribrid (indeed, reinjection of the extract of CPN2/OE484 on the UHPLC-Orbitrap ID-X Tribrid yielded identical retention times for the set of compounds). The same explanation applies to **Supplementary Figs. 10** and **11**.

a

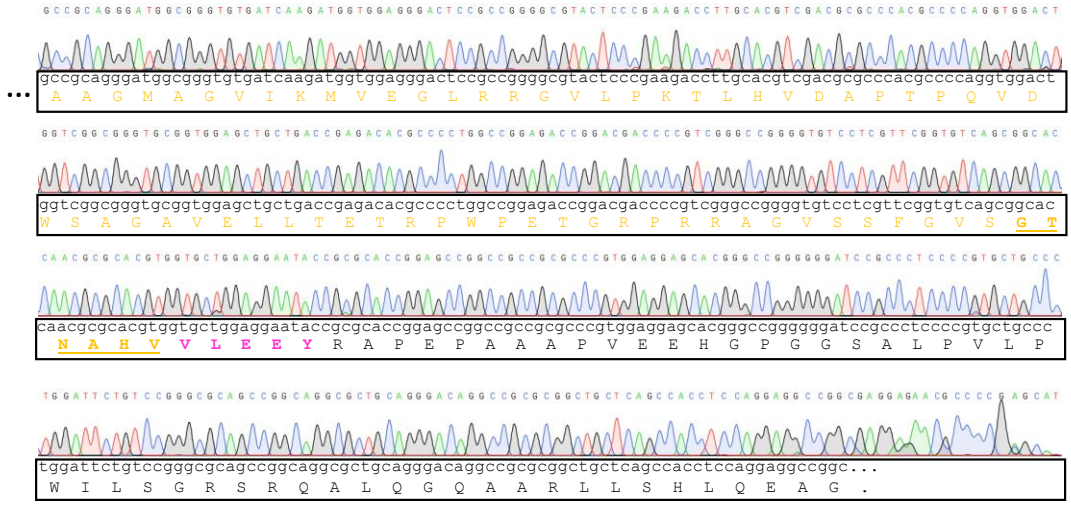


Supplementary Fig. 10 Engineering the stambomycin PKS using the alternative module definition. a Sequence alignment of KS_{14} and KS_{21} with the KS domains from the rapamycin PKS, highlighting the two alternative conserved sequence motifs that were exploited in our experiments. Motif "APNGP" (in orange) is located near the middle of the KS domains and corresponds to a recombination hot spot identified during induced evolution of the rapamycin PKS[11]. Motif "GTNAHV" (in red), at the C-terminal end of the KS domains, is a site identified empirically by genetic engineering that allows for efficient AT swapping[12].

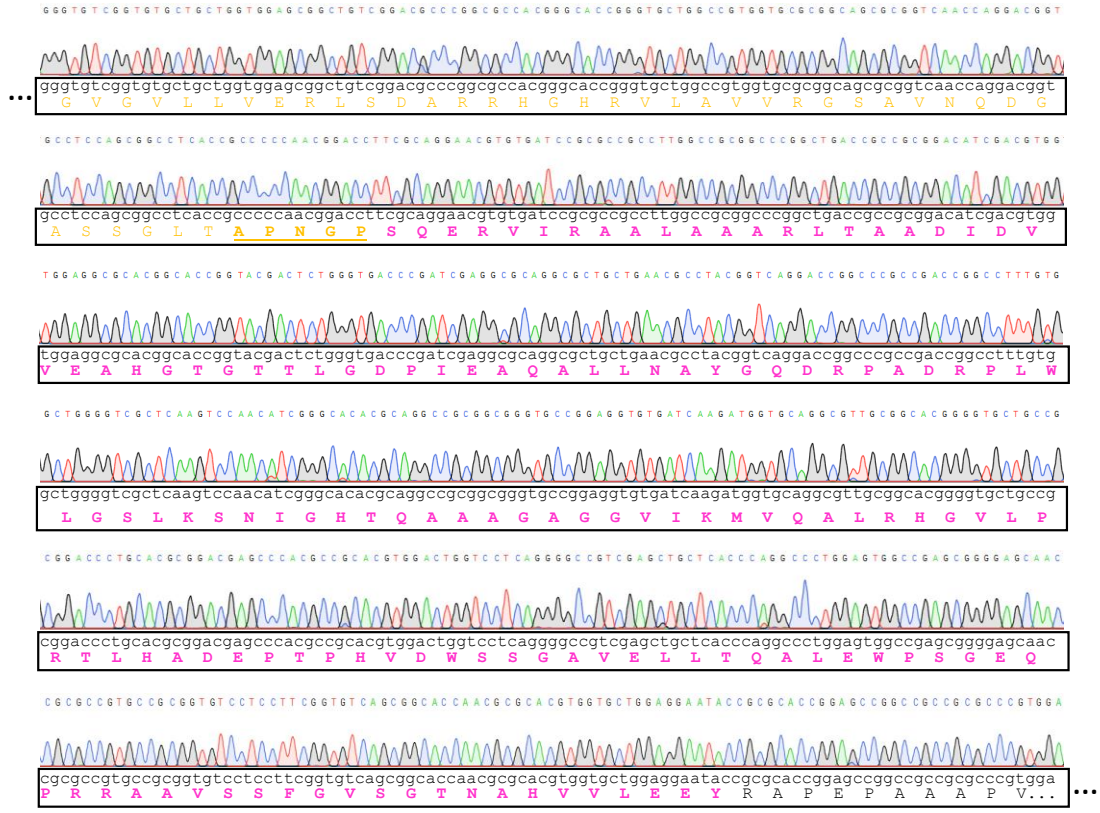
b


5′-CCCGTGCTCTTCGGTATCTCGCC-3′
5′-GGGCACGAGAAGCCATAGAGCGG-3′
crRNA11
pks4
pks5
pks8
pks9
ATCC
module13
module14
module 21
DH
ER
KR
KS
AT
ACP
HAL_hy59_S1
HAR_hy59_S1
PermE* sgRNA
pCRISPR-Cas9-crRNA11-hy59_S1
NcoI
crRNA11
CCATGGGGCGAGATACCGAAGAGCACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTACGTAGG
SnaBI
HAL_hy59_S2
HAR_hy59_S2
pCRISPR-Cas9-crRNA11-hy59_S2
1. DSB occurred within the chromosomal DNA homologous to 20 bp crRNA
2. Double crossover at HAL and HAR (repair step)
3. PCR screening of target-mutated exconjugants
4. Increase cultivation temperature to promote selection of plasmid-free progeny
pks9/KS14 swap
ATCC/hy59_S1
module 21 (KS14 swap)
KS14 swap
NDD5-KS14-KS21-AT21---
MSDTDPDQQEKLLAYLKRATTQLRDANRRVRELEDRSREPIAVIGMACRYPGGVRSPEDLWKLVESGTDALGPLPTDRGWDPAR
TSAAPRGGFVHDVGDFDAGLFGISPREAAATDPQQRQLLEASWEAVERAGIDMRTLRGGRTGVFAGVMYHDYPSVVDPEALDGY
LGTANAGSVLSGRIAYTFGLQGPAVTVDTACSSSLVALHLAAQALRAGECELALVGGVTVMSGPMMFAGFGLEDGSAADGRCKA
FAAAADGTGWGEGVGVLLVERLSDARRHGHRVLAVVRGSAVNQDGASSGLTAPNGPAQQRVIRQALASAALVPAEVDAVEAHGT
GTTLGDPIEAQALLATYGQERAEPLYLGSIKSNIGHTQAAAGMAGVIKMVEGLRRGVLPKTLHVDAPTPQVDWSAGAVELLTET
RPWPETGRPRRAGVSSFGVSGTNAHVVLEEYRAPEPAAAPVEEHGPGGSALPVLPWILSGRSRQALQGQAARLLSHLQEAGEEN
APAPMDIGFSCATTRATLEHRAVVLAEDHAGFRRGLTALAEGRETAEVIARTQDPGQGGTVF... ... ...
pks9/KS14/21
ATCC/hy59_S2
module 21 (KS14/21)
KS14/21
NDD5-KS14-KS21-AT21---
MSDTDPDQQEKLLAYLKRATTQLRDANRRVRELEDRSREPIAVIGMACRYPGGVRSPEDLWKLVESGTDALGPLPTDRGWDPAR
TSAAPRGGFVHDVGDFDAGLFGISPREAAATDPQQRQLLEASWEAVERAGIDMRTLRGGRTGVFAGVMYHDYPSVVDPEALDGY
LGTANAGSVLSGRIAYTFGLQGPAVTVDTACSSSLVALHLAAQALRAGECELALVGGVTVMSGPMMFAGFGLEDGSAADGRCKA
FAAAADGTGWGEGVGVLLVERLSDARRHGHRVLAVVRGSAVNQDGASSGLTAPNGPSQERVIRAALAAARLTAADIDVVEAHGT
GTTLGDPIEAQALLNAYGQDRPADRPLWLGSLKSNIGHTQAAAGAGGVIKMVQALRHGVLPRTLHADEPTPHVDWSSGAVELLT
QALEWPSGEQPRRAAVSSFGVSGTNAHVVLEEYRAPEPAAAPVEEHGPGGSALPVLPWILSGRSRQALQGQAARLLSHLQEAGE
ENAPAPMDIGFSCATTRATLEHRAVVLAEDHAGFRRGLTALAEGRETAEVIARTQDPGQGGTVF... ... ...

Representative sequencing result of the mutant ATCC/hy59_S1 (KS_{14} swap):

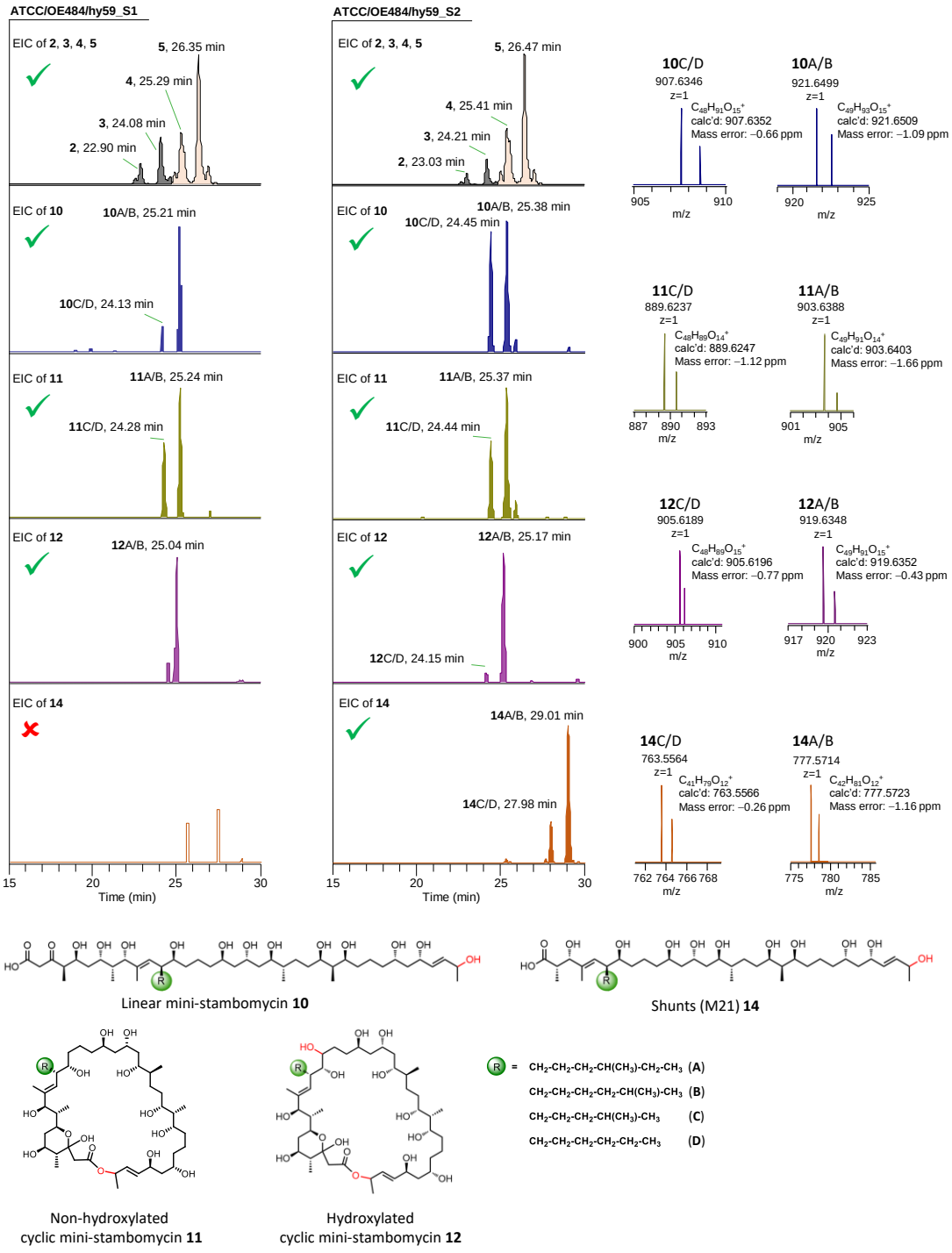


Representative sequencing result of the mutant ATCC/hy59_S2 ($KS_{14/21}$):



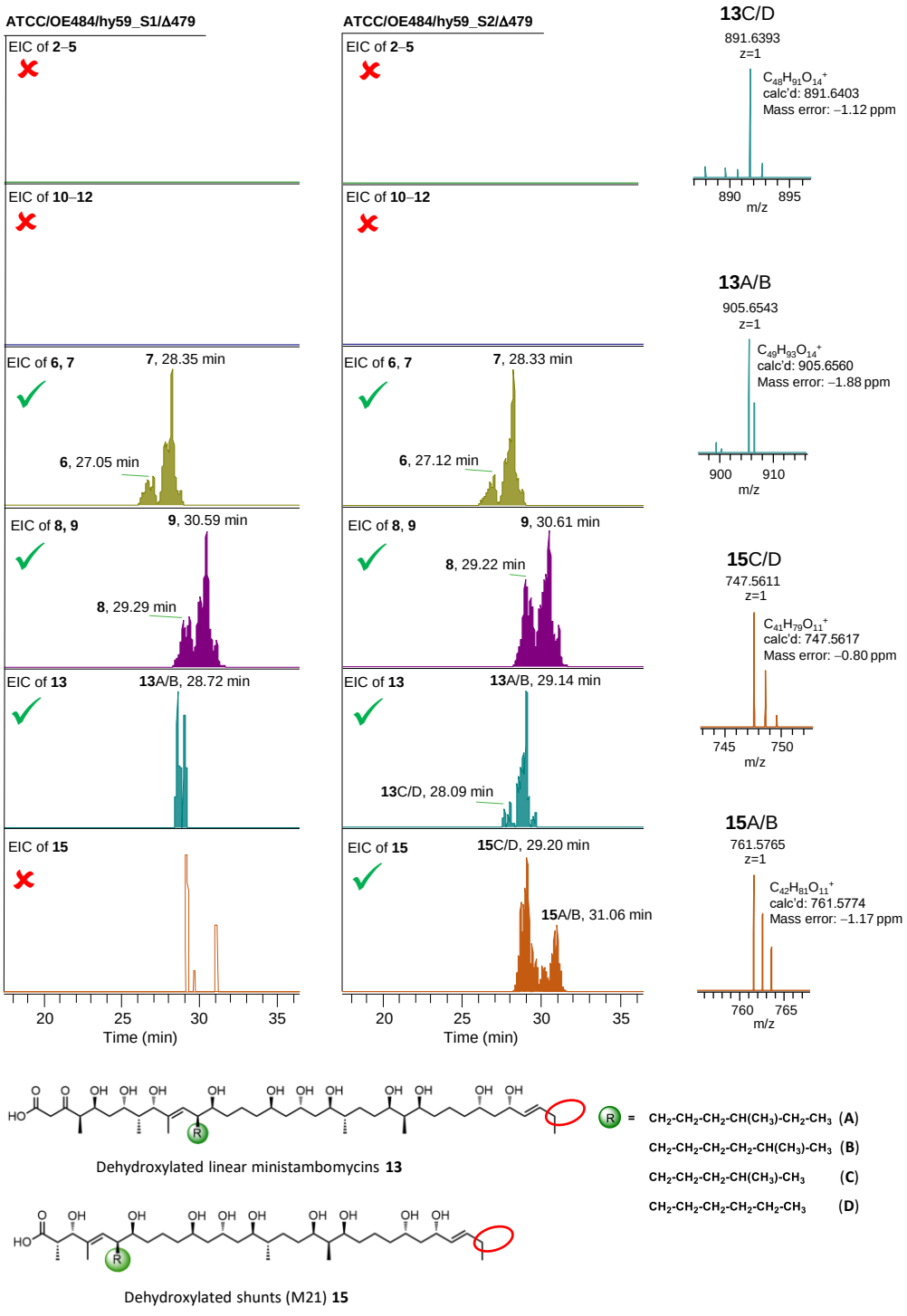
Supplementary Fig. 10 Engineering the stambomycin PKS using the alternative module definition, cont. **b** Illustration of the strategies to swap KS_{21} with KS_{14}, and to generate a hybrid $KS_{14/21}$. In the case of mutant ATCC/hy59_S1, only five residues of KS_{21} were retained, while for mutant ATCC/hy59_S2, the N-terminal docking domain of subunit 9 ($^{N}DD_{9}$) and half of KS_{21} were replaced by corresponding regions of Pks5. Also shown are representative sequencing data used to verify the accuracy of the two constructs. Amino acids in orange derive from Pks5, while those in pink and black (KS-AT linker), derive from Pks9.

c



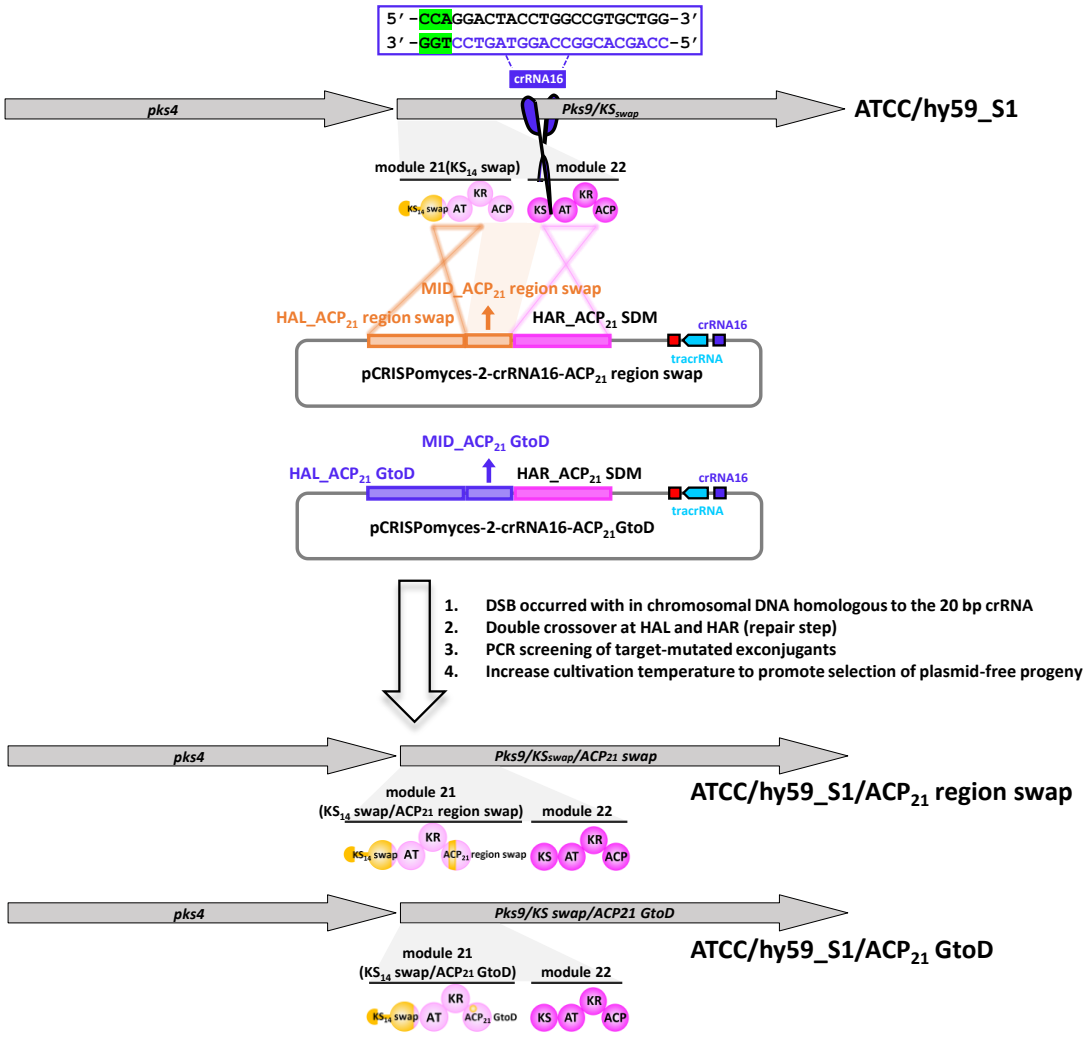
Supplementary Fig. 10 Engineering the stambomycin PKS using the alternative module definition, cont. **c** Comparative metabolic analysis of the mutants ATCC/OE484/hy59_S1 and ATCC/OE484/hy59_S2. The structures of the obtained metabolites are shown (which vary in the nature of the R group), with the two cytochrome P450-catalyzed hydroxylations indicated in red. M21 refers to the module from which the intermediate was released.

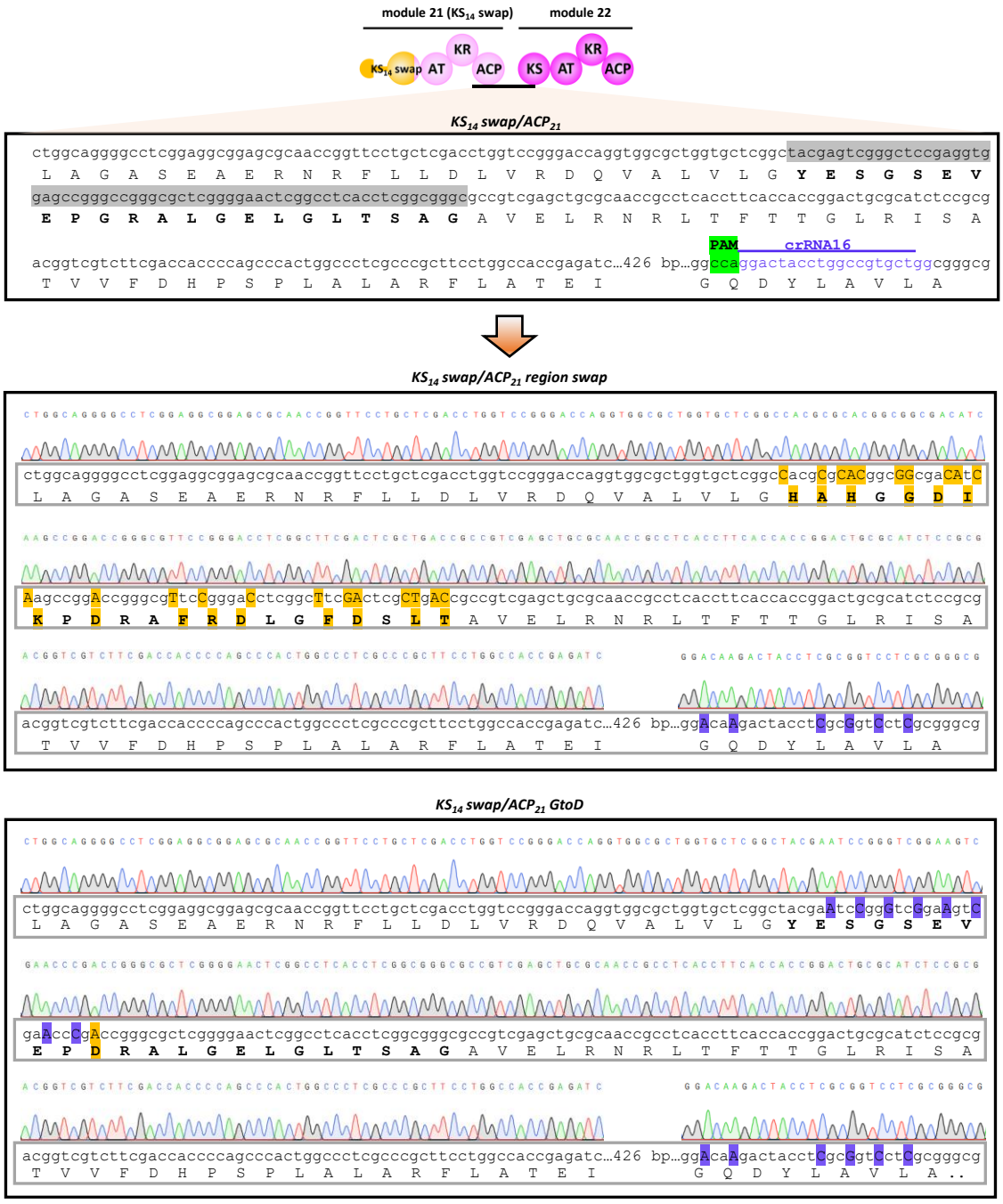
d



Supplementary Fig. 10 Engineering the stambomycin PKS using the alternative module definition, cont. **d** Inactivation of P450 hydroxylase SamR0479 in strain ATCC/OE484/hy59_S1 and ATCC/OE484/hy59_S2. As anticipated, the loss of the terminal hydroxyl group prohibited cyclization of the mini-stambomycins, resulting exclusively in production of linear, dehydroxylated **13**. Compound **15**, the product of release from the module 21 ACP, was also detected in the SamR0479 mutant of ATCC/OE484/hy59_S2.

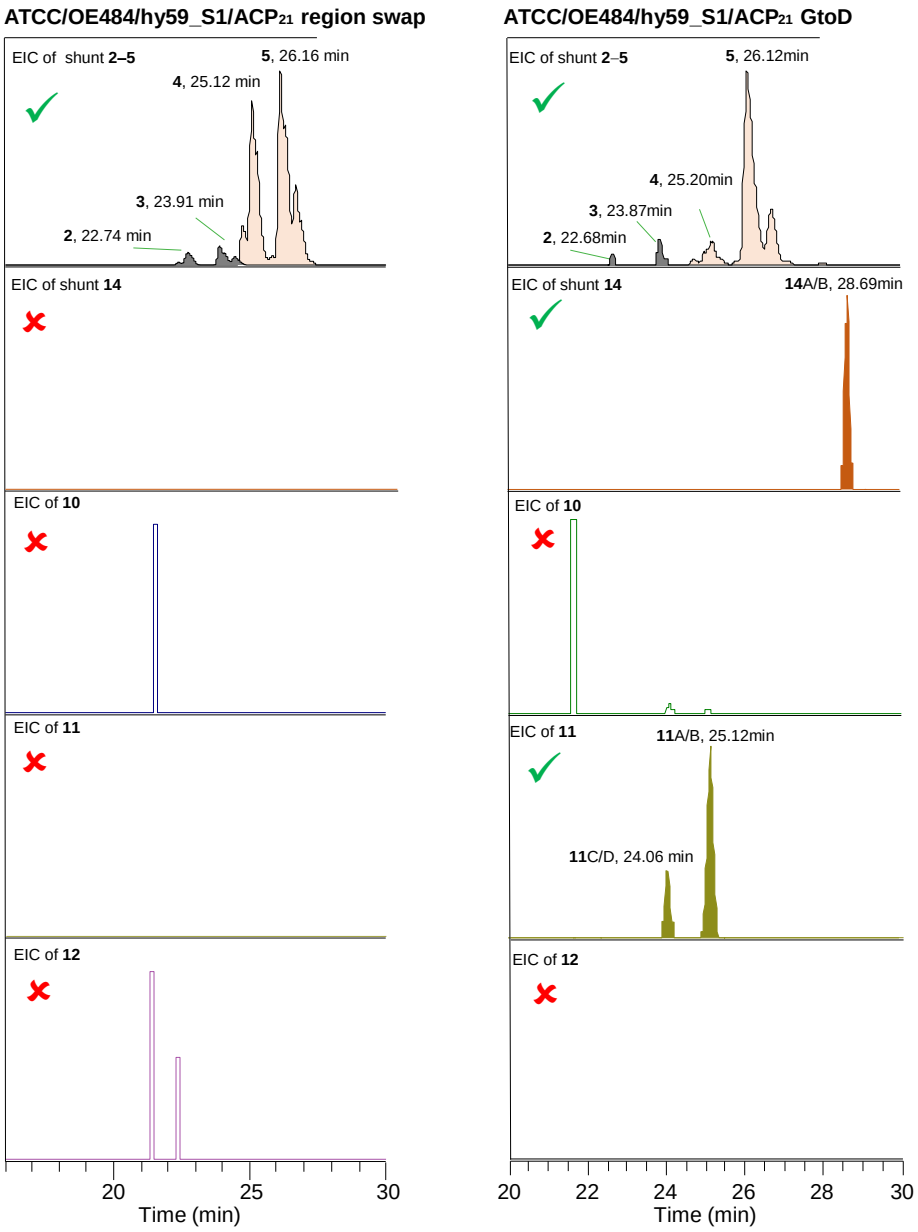
a


5′-CCAGGACTACCTGGCCGTGCTGG-3′
3′-GGTCCTGATGGACCGGCACGACC-5′
crRNA16
pks4
Pks9/KS_swap
ATCC/hy59_S1
module 21(KS14 swap)
module 22
KR
KS14 swap AT ACP
KS AT ACP
MID_ACP21 region swap
HAL_ACP21 region swap
HAR_ACP21 SDM
crRNA16
tracrRNA
pCRISPomyces-2-crRNA16-ACP21 region swap
MID_ACP21 GtoD
HAL_ACP21 GtoD
HAR_ACP21 SDM
crRNA16
tracrRNA
pCRISPomyces-2-crRNA16-ACP21GtoD
1. DSB occurred with in chromosomal DNA homologous to the 20 bp crRNA
2. Double crossover at HAL and HAR (repair step)
3. PCR screening of target-mutated exconjugants
4. Increase cultivation temperature to promote selection of plasmid-free progeny
pks4
Pks9/KSswap/ACP21 swap
ATCC/hy59_S1/ACP21 region swap
module 21
(KS14 swap/ACP21 region swap)
module 22
KS14 swap AT ACP21 region swap
KS AT ACP
pks4
Pks9/KS swap/ACP21 GtoD
ATCC/hy59_S1/ACP21 GtoD
module 21
(KS14 swap/ACP21 GtoD)
module 22
KS14 swap AT ACP21 GtoD
KS AT ACP



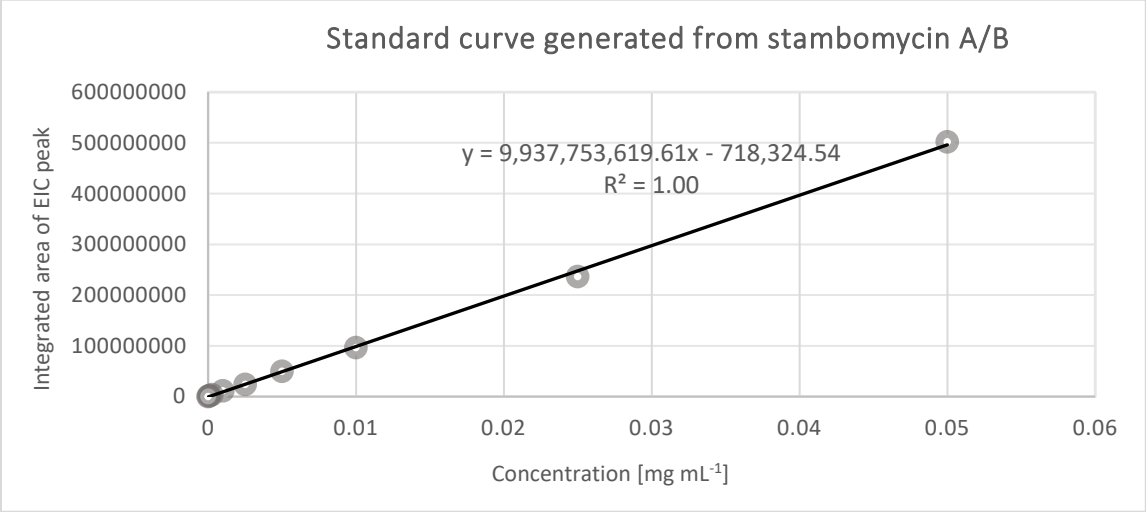
Supplementary Fig. 11 Modification of ACP_{21} in strain ATCC/hy59_S1. a Illustration of the strategy to modify ACP_{21}. Residues 31–54 within loop 1 and a portion of helix α2 (L1+H2 region, highlighted in grey) (**Supplementary Fig. 9a**) constituting the putative elongation interface on ACP_{21}[9] with its KS partner, were targeted to match those of ACP_{14} (either by exchange of the whole region (highlighted in orange) or by mutation of a single residue within a key two-amino acid recognition motif (GR → DR, as R is common for ACP_{14} and ACP_{21}; the D is highlighted in orange, while additional nucleotides highlighted in blue were also modified in order to facilitate screening for correct mutants). Successful generation of the ACP_{21} mutants was verified by DNA sequencing, as was mutation of DNA sequence corresponding to the selected crRNA16 (highlighted in blue), allowing avoidance of Cas9-catalyzed cleavage within the genome of the two resulting mutant strains.

b



Supplementary Fig. 11 Modification of ACP_{21} in strain ATCC/hy59_S1, cont. **b** LC-MS analysis of mutants ATCC/OE484/hy59_S1/ACP_{21} region swap and ATCC/OE484/hy59_S1/ACP_{21} GtoD. Only the mini-stambomycins **11** were detected in extracts of ATCC/OE484/hy59_S1/ACP_{21} GtoD, along with shunt metabolite **14**, corresponding to chain extension intermediate released from module 21.

A series of purified stambomycin A/B dilution analyzed by LC-MS.	
Concentration (mg mL^{-1})	**EIC peak area of A/B** (singly charged:1376.9392, doubly charged: 688.9736)
0.05	502106543
0.025	236663095
0.01	96152944
0.005	49462658
0.0025	23927677
0.001	11363051
0.00025	3490118
0.0001	1914743
0.00005	825977
0.00001	164391



Supplementary Fig. 12 Standard curve for quantification generated using stambomycins A/B. To generate the curve, a mixed sample of previously-purified stambomycin A/B was used as standard, with measurements made at a range of concentrations as shown in the table. Peak areas corresponding to stambomycins A/B were then calculated from the extracted ion chromatograms (using $[M+H]^+$ = 1376.9392 and $[M+2H]^{2+}$ = 688.9736). Standards with concentrations of 0.025, 0.005, and 0.001 mg mL^{-1} were measured twice, thus the areas shown in the table represent average values. This analysis yielded an essentially linear standard curve (R^2 =1.0) and equation: y = 9937753619.61x – 718324.54, which was then used to determine the yields of all stambomycin derivatives generated in this study.

Supplementary Tables

Supplementary Table 1 List of amino acid sequences of all docking domain used in this work.

Docking domain (DD)	Amino Acid sequence
$^{C}DD_4$	GPGSYADRLLAELARLEADLTRMDGDEEDRTRLAARLRALADRCQGERTRRDDNL DAASIEEVFDLLDTEFETP
$^{C}DD_4$ SDM	GPGSYADRLLAELARLEADLTRMDGDEEDRTRLAARLRALADRCQGERTRRDDNL DAASIDEVFDLIDTELETP
$^{C}DD_4$ helix swap	GPGSYADRLLAELARLEADLTRMDGDEEDRTRLAARLRALADRCQGERTRRDDNL DAADDDNLFAFIDNELGA
$^{N}DD_5$	GPGSMSDTDPDQQEKLLAYLKRATTQLRDANRRVRELEDRS
$^{C}DD_8$	GPGSYGPAALLADLDRLDAALAGTGPGDETRSRVISRLHALLARHDTADGSGLVTA AGNLDAADDDNLFAFIDNELGA
$^{N}DD_9$ Met	MPDETVGARAVSTEDKLRDYLKRVTTDLASVRQRLRAAEST
$^{N}DD_9$ Val	VSTEDKLRDYLKRVTTDLASVRQRLRAAEST

Note: Residues highlighted in yellow indicate mutated and swapped amino acids, red residues were introduced via the cloning vector pBG102, while the tyrosines indicated in green were added to facilitate monitoring during purification and measurement of protein concentration.

Supplementary Table 2 Primers used in this work.		
Oligo name	**Sequence**	**Usage**
PCR-targeting-based engineering		
9996	CGGACCAGGCGGGACGACAATCTGGACGCCGCCTCGATCGACGAGGTCTTCGACCTCATCGACACCGAGCTCGAGACGCCGTGATCGCGCGCGCTTCGTTCGGGACGAA	K7N1/4_For
9997	GTAGTCCCGCAGCTTGTCCTCGGTGGACACGGCACGTGCGCCGACGGTCTCGTCCGGCATCGCCTCAGGCTCCCATCTGCCTCTTCGTCCCGAAGCAACT	K7N1/2/3_Rev
9998	GGCGAGCGGACCAGGCGGGACGACAATCTGGACGCCGCCGATGACGATAACCTGTTCGCGTTCATCGATAACGAACTCGGAGCCTGATCGCGCGCGCTTCGTTCGGGACGAA	K7N2/5_For
A002	GTCTTCGACCTCCTCGACACCGAGTTCGAGACGCCGTGATCGCGCGCGCTTCGTTCGGGACGAA	K7N3/6_For
CDS2_amont	CCGGGTCCGTGTCGCTCATGGTCTGCTGTTCGCGTCCTCTCTGCCTCTTCGTCCCGAAGCAACT	K7N4/5/6_Rev
9999	CGACTGCTGGCGGAACTGGC	Flanking_K7N1/2/3/4/5/6 _For
A420	ATGCCGATGATGGCGATCGG	Flanking_K7N1/2/3_Rev
CK_CDS2	GTTGGCGTCGCGCAGTTG	Flanking_K7N4/5/6_Rev
110	AGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCG CCCAGTTCCGCCCATTCTC	Replacement of the chloramphenicol resistance gene of the BAC by that for kanamycin (Cm → Kana_F)
111	TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGAACC ATGAGATCCCCGCGCTGGA	Replacement of the Cm gene of the BAC by that of Kan (Cm →Kana_R)
CRISPR-Cas9-based engineering		
C976	ACGCTTGAGTGAGCCAAGGTAGAG	crRNA1_For
C977	AAACCTCTACCTTGGCTCACTCAA	crRNA1_Rev
B564	GAGACATTCTAGATGACCGCGGTGGTGCACACC	CPN1/2/3-HAL_For
B565	TTCAGGCTCTAGACCACCGCCAGCACCTCGTGC	CPN1/2/3-HAR_Rev
B571	TCACGGCGTCTCGAGCTCGGTGTCGATGAGGTCGAAGACCTCGTCGATCGAGGCGGCGTCCAGATTGTCG	CPN1-HAL_Rev
B572	CTCATCGACACCGAGCTCGAGACGCCGTGAGGCGATGCCGGACGAGACCG	CPN1-HAR_For
B569	TCAGGCTCCCAGTTCGTTGTCGATGAAGGCGAAGAGGTTGTCGTCGTCGGCGGCGTCCAGATTGTCG	CPN2-HAL_Rev
B570	GCCTTCATCGACAACGAACTGGGAGCCTGAGGCGATGCCGGACGAGACCG	CPN2-HAR_For
B573	CGGTCTCGTCCGGCATCGCCTCACGGCGTCTCGAACTC	CPN3-HAL_Rev
B574	GAGTTCGAGACGCCGTGAGGCGATGCCGGACGAGACCGTCGGCGCACGTGCCG	CPN3-HAR_For
B566	CGCCTGCGACGTGGCCGACC	Flanking_For_CPN1/2/3
B567	CGAAGGACCGTTCGGCGCGG	Flanking_Rev_CPN1/2/3
B568	GACCTCGACCGGCTCGACGC	Screening_For_CPN1/2/3
CPN4/5-HAL_For	GTGCTGGAATTCAGGTCTAGAGCGGCACCTCCTGCTCATCAGCCGCAGCGG	CPN4/5-HAL_For
CPN4/5-HAL_Rev	AGTCGGTCGGCCTCCCGTACCGGTTCGGGAGCCAGTTCGGCGCGGAG	CPN4/5-HAL_Rev

CPN4/5-HAR_For	GAAGCGGGCCACGACGCAACTGCGCGACGCCAACCGCC GCGTACGCG	CPN4/5-HAR_For
CPN4/5-HAR_Rev	TCTGCAGAATTCAGGTCTAGATGGCCGTAGGTGGCCAG CAGCGCCTGCGCC	CPN4/5-HAR_Rev
Gblock-CPN4	CGGGAGGCCGACCGACTGCTCGCCGAACTCGCGCGCCT CGAAGCCGACCTGACGCGGATGGACGGCGACGAGGAG GACCGGACCCGGCTGGCGGCCCGCCTGCGGGCCCTGGC CGACCGGTGCCAGGGCGAGCGGACCAGGCGGGACGAC AATCTGGACGCCGCCTCGATCGACGAGGTCTTCGACCTC ATCGACACCGAGCTCGAGACGCCGTGAGAGGACGCGA ACAGCAGACCATGAGCGACACGGACCCGGACCAGCAG GAGAAACTCCTGGCGTACCTGAAGCGGGCCACGACGC	Synthetic DNA fragment for construction of CPN4
Gblock-CPN5	CGGGAGGCCGACCGACTGCTCGCCGAACTCGCGCGCCT CGAAGCCGACCTGACGCGGATGGACGGCGACGAGGAG GACCGGACCCGGCTGGCGGCCCGCCTGCGGGCCCTGGC CGACCGGTGCCAGGGCGAGCGGACCAGGCGGGACGAC AATCTGGACGCCGCCGACGACGACAACCTCTTCGCCTTC ATCGACAACGAACTGGGAGCCTGAGAGGACGCGAACA GCAGACCATGAGCGACACGGACCCGGACCAGCAGGAG AAACTCCTGGCGTACCTGAAGCGGGCCACGACGCAAC	Synthetic DNA fragment for construction of CPN5
crRNA2_For	ACGCAGGCGAGCAGTTTTTCCTGC	crRNA2_For
crRNA2_Rev	AAACGCAGGAAAAACTGCTCGCCT	crRNA2_Rev
D816	GTACCGGCCTGGAGACGGCC	Flanking_For_CPN4/5
D817	TACGCCCCGGCGGAGTCCCT	Flanking_Rev_CPN4/5
D672	AGCAGGAGAAACTCCTGGCG	Screening_For_CPN4
D671	CGAGGCGCGCGAGTTCGGCG	Screening_Rev_CPN4
D670	CATCGACAACGAACTGGGAG	Screening_For_CPN5
D669	CTCCCAGTTCGTTGTCGATG	Screening_Rev_CPN5
D584	TCCGGCATGAGCACACTGAC	Walking_CPN4/5_For
D585	GGTAGTCGTGGTACATCACG	Walking_CPN4/5_Rev
D716	TTGCCGCCGGGCGTTTTTTATCTAGACCCGGCACCTGGT GCGACACCACGGAGTGC	HAL_For_Pks4+TEI
D717	GAGTCGGCGGGCGAGTTCGGTGCAGTTCGGGTGGTCG AAC	HAL-Rev_Pks4+TEI
D718	CCGAACTCGCCCGCCGACTCGCCGCCACCACGGCCGCC GACGGCCAGTCC	MID_For_Pks4+TEI
D719	CGTCGTGGCCCGCTTCAGGTTCATCGGGGCAGC TCCCCGATCCACCGGTG	MID_Rev_ Pks4+TEI
D720	ACCTGAAGCGGGCCACGACGCAACTGCGCG	HAR_For_Pks4+TEI
D721	CCTTTTTACGGTTCCTGGCCTCTAGAGCCTGCGCCTCGAT GGGATCACCCAGGGTC	HAR_Rev_Pks4+TEI
D722	CCGGCTGGTCCGGATGCCCGTACCCGGAGG	Flanking_For_Pks4+TEI
D723	GGGAGTACGCCCCGGCGGAGTCCCTCCACC	Flanking_Rev_Pks4+TEI
D724	ACGCGAGGGCGACTTCGGGGGTCG	crRNA3_For
D725	AAACCGACCCCCGAAGTCGCCCTC	crRNA3_Rev
D726	ACGCCTCCCCGATCCACCGGTGGA	crRNA4_For
D727	AAACTCCACCGGTGGATCGGGGAG	crRNA4-Rev

D728	TTGCCGCCGGGCGTTTTTTATCTAGACCGGCGTCTGGGGCGGCACCGGAGCAGCCG	HAL_For_TEI SDM step 1/2
D729	CCGATCCACCGGTGGACGGCCAGCAACTCGCCGCGACCGTAGCCGAGGAC	HAL_Rev_TEI SDM step 1
D730	GCCGTCCACCGGTGGATCGGGGAGCTGCCC	HAR_For_TEI SDM step 1
D731	CCTTTTTACGGTTCCTGGCCTCTAGATCGAGGGTGAAGTCGCGCATGTCGCCCCGG	HAR_Rev_TEI SDM step 1/2
D732	GGTCGGCGCGGGCGAGGACACCCCCTGGAC	Flanking_For_TEI SDM step 1/2
D733	AGGGTTCGATGACGATGACCCCGCCCGGTG	Flanking_Rev_TEI SDM step 1/2
D734	CCCGCGACCCCCGAAGTCGCCCTCGACTGGCAGGCCCGGG	MID_For_TEI
D735	CCGATCCAGCGGTGCACCGCGCGGGCCGTGGTGCCGGCGTGCTCGGCC	MID_Rev_TEI
D736	GCGGTGCACCGCTGGATCGGCGAGCTGCCCCGATGACCGCGGTGG	HAR_For_TEI SDM step 2
D737	GCGACTTCGGGGGTCGCGGGCAGCAACTCGCCGCGACCGTAGCCGAGGAC	HAL_Rev_TEI SDM step 2
D781	CGTACTCTTCGGGCACGCCGGTGGCGCGATCC	MID_For_TEI SDM
D779	GGATCGCGCCACCGGCGTGCCCGAAGAGTACG	MID_Rev_TEI SDM
D879	ACGCGAGCGAGACGTTGTGAGACC	crRNA7_For
D880	AAACGGTCTCACAACGTCTCGCTC	crRNA7_Rev
D788	ACGCGGGCTCGTAACGGTAGGTCT	crRNA6-For
D787	AAACAGACCTACCGTTACGAGCCC	crRNA6-Rev
D742	TTGCCGCCGGGCGTTTTTTATCTAGACGAAGGTTCGTGGAGCGAGGACCGGCCAC	HAL_For_TEII SDM step1/2
D743	GCCGCCGAGACCTACCGTTACGAGCCCGGC	HAL_Rev_TEII SDM step1
D744	TAACGGTAGGTCTCGGCGGCGCTCGCGGAGCCGCCCGCGTGCGGCAGGC	HAR_For_TEII SDM step1
D745	CCTTTTTACGGTTCCTGGCCTCTAGATGGACGACCGTACTCGCGGGCGGGCGCTGC	HAR_Rev_TEII SDM step1/2
D746	GCCGCGGAATCACATGTGGACCGCTCGATC	Flanking_For_TEII SDM
D747	GGTGAGTTGGTCCGTTCCTGGGGCATCGAC	Flanking_Rev_TEII SDM
D748	TAACGATACGTCTCGGCGGCCTTGTAGTCGGCGCGGATCGACGGCAGG	MID_For_TEII
D749	TACTACTTCCCGGTCTCACAACGTCTCGCTCCGCGCG	MID_Rev_TEII
D750	TGTGAGACCGGGAAGTAGTAGCTCGCGGAGCCGCCCGCGTGCGGCAGGCA	HAR_For_TEII SDM step2
D751	GCCGCCGAGACGTATCGTTATGAGCCGGGCCCGCCGCTGTCGACGCCCG	HAL_Rev_TEII SDM step2
D775	CGGTGGCACCCAGAGCGTGCCCGAACAACG	MID_For_TEII SDM
D773	CGTTGTTCGGGCACGCTCTGGGTGCCACCG	MID_Rev_TEII SDM
E657	CATGCCATGGGGCGAGATACCGAAGAGCACGTTTTAGAGCTAGAAATAGC	sgRNA11_For
E659	ACGCCTACGTAAAAAAAGCACCGACTCGGTGCC	sgRNA11_Rev
E656	TCGTCGAAGGCACTAGAAGGCCTCGGACCAGGCGGGACGACAATCTGGACGCC	HAL_For_hy59_S1/2
E337	GCTGACACCGAACGAGGACACCCCGGCCCG	HAL_Rev_hy59_S1
E334	AGGTCCGTTGGGGGCGGTGAGGCCGCTGG	HAL_Rev_hy59_S2
E336	GTGTCCTCGTTCGGTGTCAGCGGCACCAACGCGCACGTGGTGCTGGAGG	HAR_For_hy59_S1
E333	CTCACCGCCCCCAACGGACCTTCGCAGGAACGTGTGATCCGCGCCGCCTTG	HAR_For_hy59_S2

E655	GGTCGATCCCCGCATATAGGCCTGCGGGCTGATCTCCA GGAACAGGGTGCGGC	HAR_Rev_hy59_S1/2
E332	GCCCGCCTCGAAGCCGACCTGACGCGGATG	Flanking_For_hy59_S1/2
E331	CCGCCGCGGCCGGCGTACACCGCGGCCCAG	Flanking_Rev_ hy59_S1/2
E587	GATGATGTTCGCGGGCTTCGGCCTGGAAGAC	Walking_For_ hy59_S1/2
E588	GGTGGTGGCGCAGGAGAAGCCGATGTCCATG	Walking_Rev_ hy59_S1/2
F300	ACGCGTCCAGATTGTCGTCCCGCC	crRNA15_For
F299	AAACGGCGGGACGACAATCTGGAC	crRNA15_Rev
F298	CCGTCCGAGCGGCGGCAGGTCGTGCTCGACCTGGTCCG CGGCCAG	MID_For_ACP13 SDM
F297	GAGGTTGTCGTCGCGGCGGGTCCGCTCGCCCTGGCACC GGTCGGCCAGGG	MID_Rev_ACP13 SDM
F296	GTGCTGGAATTCAGGTCTAGACCTCGACCCGGCGGCCG AGCCCGACCCCGC	HAL_Ror_ACP13 SDM
F295	GACCTGCCGCCGCTCGGACGGGGGCACTCCGGAGAGC GCGGCCCGCAGGGC	HAL_Rev_ACP13 SDM
F294	CCCGCCGCGACGACAACCTCGACGCCGCCGACGACGAC AACCTCTTCGCC	HAR_For_ACP13 SDM
F293	TCTGCAGAATTCAGGTCTAGAGTCAGTCCGCTGGACGC GCCGTCCTGGTTG	HAR_Rev_ACP13 SDM
F292	CCGGGCAGGCCGAGCACCCC	Flanking_For_ACP13 SDM
F291	CGGCCAAGGCGGCGCGGATC	Flanking_Rev_ACP13 SDM
F301	CCCGTCCGAGCGGCGGCAGG	Screening_For_ACP13 SDM
F302	CCTGCCGCCGCTCGGACGGG	Screening_Re_ACP13 SDM
F304	CCGAACTCGCCCGCCGACTC	Walking_For_ACP13 SDM
F303	GAGTCGGCGGGCGAGTTCGG	Walking_Rev_ACP13 SDM
F325	ACGCCCAGCACGGCCAGGTAGTCC	crRNA16_For
F324	AAACGGACTACCTGGCCGTGCTGG	crRNA16_Rev
F323	GTGCTGGAATTCAGGTCTAGACCTGCGCCGCGCCGAGC CGGCCGCGCCCCG	HAL_For_ACP21 SDM
F322	CCGGTCCGGCTTGATGTCGCCGCCGTGCGCGTGGCCGA GCACCAGCGCCACCTGGTCCCGGAC	HAL_Rev_ACP21 region swap
F321	CCGGTCGGGTTCGACTTCCGACCCGGATTCGTAGCCGA GCACCAGCGCCACCTGGTCCCGGAC	HAL_Rev_ACP21 GtoD
F320	GCGACATCAAGCCGGACCGGGCGTTCCGGGACCTCGGC TTCGACTCGCTGACCGCCGTCGAGCTGCGCAACCGCCTC ACC	MID_For_ACP21 region swap
F319	CGGAAGTCGAACCCGACCGGGCGCTCGGGGAACTCGG CCTCACCTCGGCG	MID_For _ACP21 GtoD
F318	GCCGGCCACGGGGTCGCCCGCGAGGACCGCGAGGTAG TCTTGTCCGCCGGTGCCGGCGTACACCGCCGTGCG	MID_Rev_ACP21 SDM
F317	CGGGCGACCCCGTGGCCGGCGAAGGCTACC	HAR_For_ACP21 SDM
F316	TCTGCAGAATTCAGGTCTAGACGGCGGTGTCCGCGTCCA GTCCGGCGAGC	HAR_Rev_ACP21 SDM
F315	CCTCGCCGGCACCGACGAGG	Flanking_For_ACP21 SDM
F314	GTGTGGGAGCCGAGGACCAC	Flanking_Rev_ACP21 SDM
F313	CACGCGCACGGCGGCGACATCAAG	Screening_For_ACP21 region swap
F312	GAATCCGGGTCGGAAGTCGAACCC	Screening_For_ACP21 GtoD
F311	CGAGGACCGCGAGGTAGTCTTGTC	Screening_Rev_ACP21 SDM

F310	GACCTGTGGAAGCTGGTCGAG	Walking_For_ACP21 SDM
F309	CTCGACCAGCTTCCACAGGTC	Walking_Rev_ACP21 SDM
D456	CCTGGTATCTTTATAGTCCTGT	Verification of the plasmid pCRISPomyces-2_crRNA
D864	GAAACTTCTGTGAATGGCCTGTTCG	Verification of the plasmid pCRISPomyces-2_crRNA_HAs
***Docking domain expression in* E. coli**		
E539	CTGTTCCAGGGGCCCGGATCC**TAT**GCCGACCGACTGCTGGCGGAACTGG	$^{C}DD_{4}$_For
E245	CTCGAGTGCGGCCGCAAGCTTTCACGGCGTCTCGAACTCGGTGTCG	$^{C}DD_{4}$_Rev
E538	CTGTTCCAGGGGCCCGGATCC**TAT**GCCGACCGACTGCTCGCCGAACTCG	$^{C}DD_{4}$ SDM/helix swap_For
E243	CTCGAGTGCGGCCGCAAGCTTTCACGGCGTCTCGAGCTCGGTGTCG	$^{C}DD_{4}$ SDM_Rev
E242	CTCGAGTGCGGCCGCAAGCTTTCAGGCTCCCAGTTCGTTGTCGATG	$^{C}DD_{4}$ helix swap_Rev
E241	CTGTTCCAGGGGCCCGGATCCATGAGCGACACGGACCCGGACCAGC	$^{N}DD_{5}$_For
E240	CTCGAGTGCGGCCGCAAGCTTTCACGACCGGTCCTCCAGTTCGCGTACG	$^{N}DD_{5}$_Rev
E537	CTGTTCCAGGGGCCCGGATCC**TAT**GGCCCCGCCGCGCTCCTGGCCGACC	$^{C}DD_{8}$_For
E238	CTCGAGTGCGGCCGCAAGCTTTCAGGCTCCCAGTTCGTTGTCGATG	$^{C}DD_{8}$_Rev
E247	CTGTTCCAGGGGCCCGGATCCGTGTCCACCGAGGACAAGCTGCGGG	$^{N}DD_{9}$ Val_For
E486	CTGTTCCAGGGGCCCGGATCCATGCCGGACGAGACCGTCGGCGCAC	$^{N}DD_{9}$ Met_For
E248	CTCGAGTGCGGCCGCAAGCTTTCAGGTGGACTCCGCGGCACGCAGCCGC	$^{N}DD_{9}$ Val/Met_Rev

Note: 'NNN' or 'N' represent mutagenesis nucleotides; 'N' represents the modified bases to avoid cleavage by Cas9 within the genome of the obtained mutant. "NNNNNNN" represents sequence identical to that flanking the disruption cassette in the PCR-targeting system; 'TCTAGA' is the restriction recognition site of *Xba*I, 'CCATTG' of *Nco*I; 'AGGCCT' of *Stu*I;, 'GGATCC' of *Bam*HI and 'AAGCTT' of *Hin*dIII; '**TAT**' is the incorporated codon for tyrosine. Key: HAL, Homologous Arm Left; HAR, Homologous Arm Right; SDM, Site-Directed Mutagenesis.

Supplementary Table 3 Plasmids and BACs used in this work.

Name	Properties and use (resistance)	Reference
BAC1	BAC (BAA9ZA8) from the genomic library of *Streptomyces ambofaciens,* in which the chloramphenicol resistance gene (Cm^R) was replaced by a kanamycin resistance gene (Kan^R)	This work
BAC1_K7N1	BAC1/Δ*pks5-8_aac(3)IV+oriT_SDM,* mutant of BAC1 with the genes *pks5-8* replaced by an "*aatL+aac(3)IV +oriT+ aatR"* cassette and the C-terminal docking domain of Pks4 mutated by SDM (Kan^R, $Apra^R$)	This work
BAC1_K7N2	BAC1/Δ*pks5-8_aac(3)IV+oriT_helix swap,* mutant of BAC1 with the genes *pks5-8* replaced by an "*aatL+aac(3)IV+oriT +aatR"* cassette and the C-terminal docking domain of Pks4 mutated by swapping (Kan^R, $Apra^R$)	This work
BAC1_K7N3	BAC1/Δ*pks5-8_aac(3)IV+oriT,* mutant of BAC1 with the genes *pks5-8* replaced by an "*aatL+aac(3)IV+ oriT+aatR"* cassette (Kan^R, $Apra^R$)	This work
BAC1_K7N4	BAC1/*pks4_SDM_ aac(3)IV+oriT,* mutant of BAC1 with the gene *pks4* mutated by SDM within the ^{C}DD-encoding region and with incorporation of an "*aatL +aac(3)IV+oriT+ aatR"* cassette (Kan^R, $Apra^R$)	This work
BAC1_K7N5	BAC1/*pks4_helix swap_aac(3)IV+oriT,* mutant of BAC1 with the gene *pks4* mutated by swapping of the portion encoding the ^{C}DD$_4$ and with incorporation of an "*aatL+aac(3)IV+oriT +aatR"* cassette (Kan^R, $Apra^R$)	This work
BAC1_K7N6	BAC1/*pks4_helix swap_aac(3)IV+oriT,* mutant of BAC1 with incorporation of an "*aatL+aac(3)IV+oriT +aatR"* cassette at the end of *pks4* (Kan^R, $Apra^R$)	This work
BAC2	BAB11ZA7 from the genomic library of *S. ambofaciens,* in which the chloramphenicol resistance gene (Cm^R) is replaced by a kanamycin resistance gene (Kan^R)	Ref [13]
BAC2_K7Δ478	BAC2/Δ*samR0478_aadA+oriT;* mutant of BAC2 with the gene *samR0478* replaced by a *"FRT+aadA+oriT +FRT"* cassette (Kan^R, $Spec^R$)	Ref [13]
BAC2_K7Δ479	BAC2/Δ*samR0479_aadA+oriT;* mutant of BAC2 with the gene *samR0479* replaced by a *"FRT+aadA+oriT +FRT"* cassette (Kan^R, $Spec^R$)	Ref [13]
pIJ776	Origin of kanamycin resistance gene (Kan^R)	Ref [14]
pIJ778	Origin of "*FRT+aadA+oriT+FRT"* cassette ($Spec^R$)	Ref [14]
pIJ773	Origin of "*FRT+aac(3)IV+oriT+FRT*" cassette, ($Apra^R$)	Ref [14]
pSPM88T	Origin of "*aatL+aac(3)IV+oriT+aatR"* cassette used in the construction of BAC1_K7N1-N6 ($Apra^R$)	Ref [15]
pOSK1111	Conjugative plasmid for "*aatL+aac(3)IV+oriT+aatR"* cassette excision (pSAM2 genes *xis* and *int)* (Amp^R, $Thio^R$)	Ref[15]
pUWL-*oriT-flp*	Conjugative plasmid for "*FRT+aac(3)IV+oriT+FRT"* cassette excision (recombinase FLP) (Amp^R, Hyg^R)	Ref [16]
pIB139	Conjugative and integrative vector (ϕC31 *attP-int, ermEp**) ($Apra^R$)	Ref [17]

pOE484	pIB139+*samR0484* (Apra^R), regulator *samR0484* overexpression plasmid	Ref [18]
pCR^{TM}-Blunt	Subcloning vector (Kan^R)	Invitrogen
pCRISPomyces-2	CRISPR-Cas9 cloning vector (Apra^R)	Ref [4]
pCRISPomyces-2-crRNA1-CPN1	Engineering of mutant CPN1 (Apra^R)	This work
pCRISPomyces-2-crRNA1-CPN2	Engineering of mutant CPN2 (Apra^R)	This work
pCRISPomyces-2-crRNA1-CPN3	Engineering of mutant CPN3 (Apra^R)	This work
pCRISPomyces-2-crRNA2-CPN4	Engineering of mutant CPN4 (Apra^R)	This work
pCRISPomyces-2-crRNA2-CPN5	Engineering of mutant CPN5 (Apra^R)	This work
pCRISPomyces-2-crRNA2-Pks4+TEI	Engineering of TE fused to gene *pks4* (Apra^R)	This work
pCRISPomyces-2-crRNA3-TEI SDM_S1	Inactivation of TE step 1 (S1) (Apra^R)	This work
pCRISPomyces-2-crRNA4-TEI SDM_S2	Inactivation of TE step 2 (S2) (Apra^R)	This work
pCRISPomyces-2-crRNA7-TEII SDM_S1	Inactivation of TEII (Apra^R)	This work
pCRISPomyces-2-crRNA6-TEII SDM_S2	Inactivation of TEII (Apra^R)	This work
pCRISPomyces-2-crRNA15-ACP_{13} SDM	Engineering of mutant CPN2/OE484/ACP_{13} SDM (Apra^R)	This work
pCRISPomyces-2-crRNA16-ACP_{21} region swap	Engineering of mutant ATCC/OE484/hy59_S1/ ACP_{21} region swap (Apra^R)	This work
pCRISPomyces-2-crRNA16-ACP_{21} GtoD	Engineering of mutant ATCC/OE484/hy59_S1/ ACP_{21} GtoD (Apra^R)	This work
pCRISPR-Cas9	CRISPR-Cas9 cloning vector (Apra^R, Thio^R)	Ref [5]
pCRISPR-Cas9-crRNA11-hy59_S1	Engineering of mutant ATCC/OE484/hy59_S1 (Apra^R)	This work
pCRISPR-Cas9-crRNA11-hy59_S2	Engineering of mutant ATCC/OE484/hy59_S2 (Apra^R)	This work
pBG102	Recombinant protein expression plasmid, a derivative of pET27b, containing an N-terminal His_6-SUMO-tag, which can be cleaved by H3C protease (Kan^R)	Vanderbilt University
pBG102_$^{C}DD_4$	Overexpression of $^{C}DD_4$ wild type (Kan^R)	This work
pBG102_$^{C}DD_4$ SDM	Overexpression of $^{C}DD_4$ SDM (Kan^R)	This work
pBG102_$^{C}DD_4$ helix swap	Overexpression of $^{C}DD_4$ helix swap (Kan^R)	This work
pBG102_$^{N}DD_5$	Overexpression of $^{N}DD_5$ (Kan^R)	This work
pBG102_$^{C}DD_8$	Overexpression of $^{C}DD_8$ (Kan^R)	This work
pBG102_$^{N}DD_9$ Val	Overexpression of $^{N}DD_9$ Val (Kan^R)	This work
pBG102_$^{N}DD_9$ Met	Overexpression of $^{N}DD_9$ Met (Kan^R)	This work

Supplementary Table 4 Strains used in this work.

Strains/mutants	Description and use (resistance)	Reference
S. ambofaciens		
ATCC	*S. ambofaciens* ATCC 23877 wild type strain	Ref [19]
ATCC/pIB139	ATCC containing the empty vector pIB139 (Apra^R)	Ref [18]
ATCC/OE484	Stambomycins producing strain; ATCC containing the vector pOE484 (Apra^R)	Ref [18]
K7N1/pIB139	ATCC/pIB139/*Δpsk5-8_scar_SDM;* PCR-targeting mutant of ATCC/pIB139 with deletion of intervening genes *pks5-8,* and C-terminal docking domain of Pks4 mutated by SDM (Apra^R)	This work
K7N1/OE484	ATCC/pOE484/*Δpsk5-8_scar_SDM;* PCR-targeting mutant of ATCC/pOE484 with deletion of intervening genes *pks5-8*, and C-terminal docking domain of Pks4 mutated by SDM (Apra^R)	This work
K7N2/pIB139	ATCC/pIB139/*Δpsk5-8_scar_helix swap;* PCR-targeting mutant of ATCC/pIB139 with deletion of intervening genes *pks5-8*, and C-terminal docking domain of Pks4 mutated by swapping (Apra^R)	This work
K7N2/OE484	ATCC/pOE484/*Δpsk5-8_scar_helix swap;* PCR-targeting mutant of ATCC/pOE484 with deletion of intervening genes *pks5-8*, and C-terminal docking domain of Pks4 mutated by swapping (Apra^R)	This work
K7N3/pIB139	ATCC/pIB139/*Δpsk5-8_scar;* PCR-targeting mutant of ATCC/pIB139 with deletion of intervening genes *pks5-8* (no modification of the C-terminal docking domain of Pks4) (Apra^R)	This work
K7N3/OE484	ATCC/pOE484/*Δpsk5-8_scar;* PCR-targeting mutant of ATCC/pOE484 with deletion of intervening genes *pks5-8* (no modification of the C-terminal docking domain of Pks4) (Apra^R)	This work
K7N4/pIB139	ATCC/pIB139/*scar_SDM;* PCR-targeting mutant of ATCC/pIB139 with C-terminal docking domain of Pks4 mutated by SDM (Apra^R)	This work
K7N4/OE484	ATCC/pOE484/*scar_SDM;* PCR-targeting mutant of ATCC/pIB139 with C-terminal docking domain of Pks4 mutated by SDM (Apra^R)	This work
K7N5/pIB139	ATCC/pIB139/*scar_helix swap;* PCR-targeting mutant of ATCC/pIB139 with C-terminal docking domain of Pks4 mutated by swapping (Apra^R)	This work
K7N5/OE484	ATCC/pOE484/*scar_ helix swap;* PCR-targeting mutant of ATCC/pOE484 with C-terminal docking domain of Pks4 mutated by swapping (Apra^R)	This work
K7N6/pIB139	ATCC/pOE484/*scar;* PCR-targeting mutant of ATCC/pOE484 with a scar inserted (no modification of the C-terminal docking domain of Pks4) (Apra^R)	This work
K7N6/OE484	ATCC/pOE484/*scar;* PCR-targeting mutant of ATCC/pOE484 with a scar inserted (no modification of the C-terminal docking domain of Pks4) (Apra^R)	This work
CPN1/pIB139	ATCC/pIB139/*Δpsk5-8_SDM;* CRISPR-Cas9 mutant of ATCC/pIB139 with deletion of intervening genes *pks5-8,* and C-terminal docking domain of Pks4 mutated by SDM (Apra^R)	This work

CPN1/OE484	ATCC/pOE484/Δ*psk5-8_SDM;* CRISPR-Cas9 mutant of ATCC/pOE484 with deletion of intervening genes *pks5-8,* and C-terminal docking domain of Pks4 mutated by SDM ($Apra^R$)	This work
CPN2/pIB139	ATCC/pIB139/Δ*psk5-8_helix swap;* CRISPR-Cas9 mutant of ATCC/pIB139 with deletion of intervening genes *pks5-8,* and C-terminal docking domain of Pks4 mutated by swapping ($Apra^R$)	This work
CPN2/OE484	ATCC/pOE484/Δ*psk5-8_helix swap;* CRISPR-Cas9 mutant of ATCC/pOE484 with deletion of intervening genes *pks5-8,* and C-terminal docking domain of Pks4 mutated by swapping ($Apra^R$)	This work
CPN3/pIB139	ATCC/pIB139/Δ*psk5-8;* CRISPR-Cas9 mutant of ATCC/pIB139 with deletion of intervening genes *pks5-8* (no modification of the C-terminal docking domain of Pks4) ($Apra^R$)	This work
CPN3/OE484	ATCC/pOE484/Δ*psk5-8;* CRISPR-Cas9 mutant of ATCC/pOE484 with deletion of intervening genes *pks5-8* (no modification of the C-terminal docking domain of Pks4) ($Apra^R$)	This work
CPN4/pIB139	ATCC/pIB139/*SDM;* CRISPR-Cas9 mutant of ATCC/pIB139 with C-terminal docking domain of Pks4 mutated by SDM ($Apra^R$)	This work
CPN4/OE484	ATCC/pOE484/*SDM;* CRISPR-Cas9 mutant of ATCC/pOE484 with C-terminal docking domain of Pks4 mutated by SDM ($Apra^R$)	This work
CPN5/pIB139	ATCC/pIB139/*helix swap;* CRISPR-Cas9 mutant of ATCC/pIB139 with C-terminal docking domain of Pks4 mutated by swapping ($Apra^R$)	This work
CPN5/OE484	ATCC/pOE484/*helix swap;* CRISPR-Cas9 mutant of ATCC/pOE484 with C-terminal docking domain of Pks4 mutated by swapping ($Apra^R$)	This work
ATCC/OE484/Pks4+TEI	CRISPR-Cas9 mutant of ATCC/OE484 with gene encoding TE domain directly fused to the C-termini of *pks4* ($Apra^R$)	This work
CPN2/OE484/TEI SDM	CRISPR-Cas9 mutant of CPN2/OE484 with TE inactivated by SDM of the catalytic Ser (→ Ala) ($Apra^R$)	This work
CPN2/OE484/TEII SDM	CRISPR-Cas9 mutant of CPN2/OE484 with TEII (SamR0484) inactivated by SDM of the catalytic Ser (→ Ala) ($Apra^R$)	This work
CPN2/OE484/Δ478	CPN2/OE484/Δ*samR0478_aadA+oriT;* PCR-targeting mutant of CPN2/OE484 with the gene s*amR0478* replaced by a "*FRT+aadA+ oriT+FRT"* cassette ($Spec^R$, $Apra^R$)	This work
CPN2/OE484/Δ479	CPN2/OE484/Δ*samR0479_aadA+oriT;* PCR-targeting mutant of CPN2/OE484 with the gene s*amR0479* replaced by a "*FRT+aadA+ oriT+FRT"* cassette ($Spec^R$, $Apra^R$)	This work
ATCC/OE484/hy59_S1	CRISPR-Cas9 mutant of ATCC/OE484 with a portion of the *pks9* gene directly fused to *pks5* at the end of the region encoding its first KS domain (KS_{14} swap) ($Apra^R$)	This work
ATCC/OE484/hy59_S2	CRISPR-Cas9 mutant of ATCC/OE484 with a portion of gene *pks9* directly fused to *pks5* in the region encoding the middle of its first KS domain ($KS_{14/21}$) ($Apra^R$)	This work
CPN2/OE484/ACP_{13} SDM	CRISPR-Cas9 mutant of CPN2/OE484 in which helix α1 of ACP_{13} was mutated by SDM ($Apra^R$)	This work
ATCC/OE484/hy59_S1/ACP$_{21}$ region swap	CRISPR-Cas9 mutant of ATCC/OE484/hy59_S1 in which the loop 1 and helix α2 regions of ACP_{21} were mutated ($Apra^R$)	This work
ATCC/OE484/hy59_S1/ACP$_{21}$ GtoD	CRISPR-Cas9 mutant of ATCC/OE484/hy59_S1 in which the residue Gly on loop 1 of ACP_{21} was mutated (→ Asp) ($Apra^R$)	This work

ATCC/OE484/hy59_S1/Δ479	ATCC/OE484/hy59_S1/Δ*samR0479_aadA+oriT;* PCR-targeting mutant of ATCC/OE484/hy59_S1 with the gene s*amR0479* replaced by a "*FRT+aadA+oriT+FRT"* cassette (Spec^R, Apra^R)	This work
ATCC/OE484/hy59_S2/Δ479	ATCC/OE484/hy59_S2/Δ*samR0479_aadA+oriT;* PCR-targeting mutant of ATCC/OE484/hy59_S2 with the gene s*amR0479* replaced by a "*FRT+aadA+oriT+FRT"* cassette (Spec^R, Apra^R)	This work
E. coli		
DH5α	General cloning strain	Ref [20]
BW25113/pKD20	PCR-targeting mutagenesis strain containing a λ-RED recombination plasmid pKD20 (Amp^R)	Ref [21]
ET12567/pUZ8002	Non-methylating strain containing a mobilization plasmid for conjugation with *Streptomyces* (Kan^R, Cm^R)	Ref [22]
BL21	BL21 (DE3), a widely used T7 expression strain	Novagen
R2	Rosetta^{TM} 2 (DE3), a BL21 (DE3)-derived strain, designed to enhance the expression of proteins that contain codons rarely used in *E. coli* (Cm^R)	Novagen
BL21/pBG102_^{C}DD8	Strain for ^{C}DD8 overexpression (Kan^R)	This work
BL21/pBG102_^{N}DD5	Strain for ^{N}DD5 overexpression (Kan^R)	This work
BL21/pBG102_^{N}DD9 Val	Strain for ^{N}DD9 Val overexpression (Kan^R)	This work
BL21/pBG102_^{N}DD9 Met	Strain for ^{N}DD9 Met overexpression (Kan^R)	This work
R2/pBG102_^{C}DD4	Strain for ^{C}DD4 overexpression (Kan^R, Cm^R)	This work
R2/pBG102_^{C}DD4 SDM	Strain for ^{C}DD4_SDM overexpression (Kan^R, Cm^R)	This work
R2/pBG102_^{C}DD4 helix swap	Strain for ^{C}DD4_swap overexpression (Kan^R, Cm^R)	This work

Supplementary Table 5 Quantification of wild type stambomycins and derivatives form various mutants.

Wild type and DD engineering strains producing stambomycins

Strains (culture volume)	Integrated area of stambomycins **1** A, B, C, D (EI = 1362.9236, 681.9659, 1376.9392, 688.9736)	Mass of cells (g)	Volume of final extract in MeOH (μL)	Calculated titer (mg L^{-1})	Average ± deviation (wild type level, %)
20190130-9-WTOE (50ml)	4672442552	3.8	300	28.2	30 ± 2 (100%)
20190913-13-WTOE (50ml)	6621933694	3.2	240	32	
20190206-4-k7n6 (50ml)	4911596030	2.9	200	19.8	22 ± 3 (73%)
20190206-6-k7n6 (50ml)	5596413718	3	210	23.7	
20180719-2-k7n4 (50ml)	1354939122	3.2	225	6.1	5 ± 1 (18%)
20190206-4-k7n4 (50ml)	1308555295	2.6	180	4.7	
20181211-8-cpn4 (50ml)	2173114763	2.7	190	8.3	7 ± 2 (23%)
20181211-9-cpn4 (50ml)	1730753566	2.3	160	5.6	
20181211-11-cpn5 (50ml)	848360698	2.8	195	3.3	4 ± 1 (14%)
20181211-12-cpn5 (50ml)	1389831827	2.7	190	5.3	

DD engineering strains producing shunt products

Strains (culture volume)	Integrated area of peaks corresponding to shunt products				Mass of cells (g)	Volume of final extract in MeOH (μL)	Calculated titers (μg L^{-1}) of 2, 3, 4 and 5, respectively (wild type level, %)
	2 (665.4825 [M+H])	**3** (679.4987 [M+H])	**4** (705.5134 [M+H], 727.4972 [M+Na])	**5** (719.5291 [M+H], 741.5129 [M+Na])			
20190226-9-k7n1/OE (50ml)	48227	116879	239428	1992226	3.2	220	
20190405-22-k7n1/OE (50ml)	30888	76662	140487	924533	2.2	170	
20190206-1-k7n1/OE (50ml)	198822	302791	234494	1493742	3	210	
average	92646	165444	204803	1470167		200	0.66, 0.95, 1.1,6.2 (0.030%)
20190226-5-k7n2/OE (50ml)	68169	206607	1365494	1831999	2.4	170	
20190405-23-k7n2/OE (50ml)	5585	13638	526592	1296303	2.8	200	
average	36877	110123	946043	1564151		185	0.40,0.67, 3.8, 6.1 (0.037%)
20190206-3-k7n3/OE (50ml)	42023	186837	1124713	2296192	2	140	
20190405-24-k7n3/OE (50ml)	n.d.	n.d.	235449	265668	3	210	
average	42023	186837	680081	1280930		175	0.39, 0.91, 2.6,4.8 (0.029%)
20180719-3-k7n5/OE (50ml)	67382	81574	83076	1221210	2.5	180	
20190206-5-k7n5/OE (50ml)	34500	93184	692770	1962094	1.6	120	
20190405-25-k7n5/OE (50ml)	23563	45805	801063	1377565	3	210	
average	41815	73521	525636	1520290		170	0.38, 0.49, 2.0, 5.4 (0.028%)
20190206-7-cpn1/OE (50ml)	51402	197611	1139676	1799995	2	140	
20190206-8-cpn1/OE (50ml)	58567	158339	979740	2142216	1.7	130	
20190405-26-cpn1/OE (50ml)	7333	21868	871163	1337523	3	210	
average	39101	125939	996860	1759911		160	0.35, 0.63, 3.4, 5.9 (0.034%)
20180719-8-cpn2/OE (50ml)	23569	322174	810205	1038139	2.8	200	
20190206-9-cpn2/OE (50ml) ✓	73450	169845	1062465	2364460	2.4	150	
20190206-10-cpn2/OE (50ml)	55026	132822	933593	2287611	2.5	150	
20190405-27-cpn2/OE (50ml)	3405	10586	780584	1087050	3	210	
average	38863	158857	896712	1694315		178	0.39, 0.82, 3.5, 6.3 (0.037%)
20190130-5-Pks4+TEI/OE (50ml)	622957	1920826	22589795	43419150	3.8	150	
20190130-6-Pks4+TEI/OE (50ml)	740404	1864042	20950108	41939099	4	150	
20190130-17-Pks4+TEI/OE (50ml)	345519	944340	16606807	30752675	2.7	200	
20190130-18-Pks4+TEI/OE (50ml)	216270	587675	11314880	20453416	2.6	200	
average	481288	766008	17865398	34141085		175	1.9, 2.9, 63, 121 (0.63%)
20190405-37-cpn2/OE/TEI SDM (50ml*2)	2659	3581	338990	451109	4.8	350	0.26, 0.26, 1.4, 1.8 (0.012%)
20190529-7-cpn2/OE/TEII SDM (50ml)	3928	27230	479426	1070306	2.65	190	
20190529-8-cpn2/OE/TEII SDM (50ml)	4232	34540	620282	1353119	2.7	200	
average	4080	30885	549854	1211713		195	0.29, 0.40, 2.4, 5.0 (0.027%)
20190405-6-cpn2/OE/Δ478 (50ml)	19451	67231	823021	1458241	3	210	
20190405-7-cpn2/OE/Δ478 (50ml)	19789	47188	792059	1385322	2.6	190	
average	19620	57210	807540	1421782		200	0.36, 0.51, 3.5, 6.0 (0.035%)

DD engineering strainCPN2/OE484 in which SamR0479 is inactivated

Strains (culture volume)	Integrated area of dehydroxylated shunt products				Mass of cells (g)	Volume of final extract in MeOH (μL)	Calculated titers (μg L^{-1}) of 6, 7, 8 and 9, respectively
	6 (649.4889 [M+H])	**7** (663.5051 [M+H])	**8** (689.5198 [M+H], 711.5023 [M+Na])	**9** (703.5355 [M+H], 725.5180 [M+Na])			
20190130-23-cpn2/OE/Δ479 (50ml)	14429	43625	667345	1236812	2.8	200	
20190130-24-cpn2/OE/Δ479 (50ml)	15398	40334	335929	810120	3	210	
average	14914	41980	501637	1023466		205	0.35, 0.46, 2.4, 4.5 (0.026%)

Strains (culture volume)	Integrated peak areas (and calculated yield µg L^{-1}) of shunt products						Linear ministambomycin 10		Non-hydroxylated cyclic ministambomycin 11		Hydroxylated cyclic ministambomycin 12		Mass of cells (g)	Volume of final extract in MeOH (µL)
	2 (665.4825 [M+H]) (**M12**)	**3** (679.4987 [M+H]) (**M12**)	**4** (705.5134 [M+H], 727.4972 [M+Na]) (**M13**)	**5** (719.5291 [M+H], 741.5129 [M+Na]) (**M13**)	**14**C/D (763.5566 [M+H]) (**M21**)	**14**A/B (777.5713 [M+H]) (**M21**)	**10**C/D (907.6352 [M+H])	**10**A/B (921.6509 [M+H])	**11**C/D (889.6247 [M+H])	**11**A/B (903.6403 [M+H])	**12**C/D (905.6196 [M+H])	**12**A/B (919.6352 [M+H])		
20200619-KS_{14} swap /OE-1 (50ml)	6293693 (5.2)	5052020 (4.3)	34501397 (26)	25414414 (19)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	5	370
20200619- KS_{14} swap /OE-2 (50ml) ✓	42879664 (29)	96663771 (66)	34313367 (232)	584008582 (394)	n.d.	n.d.	84543 (0.53)	420957 (0.76)	1206749 (1.3)	2074971 (1.9)	n.d.	725054 (0.97)	4.5	335
20200619-KS_{14} swap /OE-3 (50ml)	7089726 (6.1)	18341831 (15)	53007611 (42)	75394172 (60)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	5.2	390
20200619-KS_{14} swap /OE-4 (50ml*3)	30757896 (15)	400063135 (188)	215411738 (102)	220952290 (104)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	9.5	700
20200619-KS_{14} swap /OE-5 (50ml*3)	6247919 (3.6)	5054689 (3.0)	33375699 (18)	25597428 (14)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	10.3	770
20200619-$KS_{14/21}$ /OE-1 (50ml)	47039781 (29)	85889429 (52)	315910063 (191)	406463091 (246)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.1	300
20200619-$KS_{14/21}$ /OE-2 (50ml)	18613021 (15)	40766899 (31)	140489502 (107)	198526766 (150)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	5	375
20200619-$KS_{14/21}$ /OE-3 (50ml) ✓	107314758 65)	242618264 (147)	770881040 (466)	1120574541 (677)	n.d.	766578 (0.89)	2018597 (1.6)	2844624 (2.1)	5031092 (3.5)	9617988 (6.2)	262071 (0.58)	4187448 (3.0)	4.1	300
20200619-$KS_{14/21}$ /OE-4 (50ml*3)	91405522 (45)	123907364 (61)	600891652 (290)	710632670 (343)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	9.6	720
20200619-$KS_{14/21}$ /OE-5 (50ml*2)	87894343 (44)	139795773 (69)	682188694 (337)	829685866 (409)	3641918 (2.1)	13077739 (6.8)	1570620 (1.1)	1046799 (0.86)	4674276 (2.6)	4526109 (2.6)	1605608 (1.1)	2979577 (1.8)	6.6	490
CPN2/OE484/ACP_{13} SDM														
20201120-CPN2/ACP_{13} SDM-OE1(50mL)	12640195 (6.6)	32997824 (17)	308796712 (153)	809571533 (400)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	3.5	245
20201120-CPN2/ACP_{13} SDM-OE2 (50mL) ✓	6362591 (4.0)	17482021 (10)	235144379 (133)	670480570 (378)	n.d.	n.d.	n.d.	n.d.	n.d.	179421 (0.50)	n.d.	n.d.	4	280
20201130-CPN2/ACP_{13} SDM-OE3 (50mL*3)	6704419 (3.5)	18944889 (9.2)	135610351 (64)	380365058 (179)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	10	700
20201130-CPN2/ACP_{13} SDM-OE4 (50mL*3)	9101131 (3.9)	26836822 (11)	127172374 (52)	358181384 (144)	n.d.	n.d.	n.d.	n.d.	n.d.	218373 (0.37)	n.d.	n.d.	8.5	600
20201130-CPN2/ACP_{13} SDM-OE5 (50mL*3)	10548306 (6.0)	25926832 (14)	146224041 (78)	415504574 (222)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	10.5	800
ATCC/hy59_S1 (KS_{14} swap) + ACP_{21} region swap														
20201120- KS_{14} swap /ACP21 swap-OE1 (50mL) ✓	13839888 (9.2)	22759248 (15)	430220249 (273)	555230058 (352)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.5	315
20201120- KS_{14} swap /ACP21 swap-OE2 (50mL)	10647439 (6.4)	10943471 (6.6)	292119102 (165)	241546700 (137)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4	280
20201130- KS_{14} swap /ACP21 swap-OE3 (50mL*3)	11696575 (7.5)	21720871 (14)	264715440 (160)	392408919 (237)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	13	900
20201130- KS_{swap} /ACP21 swap-OE4 (50mL*3)	6352427 (4.0)	8198604 (5.1)	147643760 (85)	199699615 (114)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	12.5	850
ATCC/hy59_S1 (KS_{14} swap) /ACP_{21} GtoD														
20201218-KS_{14} swap /ACP21 GtoD-OE1 (50mL)	5790815 (2.6)	9526123 (4.1)	51985855 (21)	258802864 (104)	n.d.	n.d.	n.d.	n.d.	205370 (0.37)	673895 (0.55)	n.d.	n.d.	2.8	200
20201218-KS_{14} swap /ACP21 GtoD-OE2 (50mL) ✓	1873317 (1.0)	6897912 (3.1)	37197787 (15)	274426649 (111)	n.d.	285110 (0.40)	n.d.	n.d.	154018 (0.34)	958112 (0.67)	n.d.	n.d.	2.6	200
20201218-KS_{14} swap /ACP21 GtoD-OE3 (50mL)	3979784 (1.9)	14314951 (6.0)	67053419 (27)	215967586 (87)	n.d.	n.d.	n.d.	n.d.	n.d.	492423 (0.48)	n.d.	n.d.	2.2	200
20201218-KS_{14} swap /ACP21 GtoD-OE4 (50mL)	5173540 (2.4)	17037931 (7.1)	64276652 (26)	189687875 (77)	n.d.	n.d.	n.d.	n.d.	161170 (0.35)	433372 (0.45)	n.d.	n.d.	3.0	200
20201218-KS_{14} swap /ACP21 GtoD-OE5 (50mL)	2120956 (1.1)	8289793 (3.6)	46422798 (19)	191692136 (77)	n.d.	248873 (0.38)	n.d.	n.d.	n.d.	771030 (0.59)	n.d.	n.d.	2.4	200
20201218-KS_{14} swap /ACP21 GtoD-OE6 (50mL)	5742932 (2.6)	10447012 (4.5)	82037304 (33)	244427835 (99)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.4	200

ATCC/OE484/hy59_S1 (KS_{14} swap)/Δ479 and ATCC/OE484/hy59_S2 ($KS_{14/21}$)/Δ479										
	Integrated peak areas of dehydroxylated shunt products						Linear dehydroxylated ministambomycin 13		Mass of cells (g)	Volume of final extract in MeOH (µL)
	6 (649.4889 [M+H]) (**M12**)	**7** (663.5051 [M+H]) (**M12**)	**8** (689.5198 [M+H], 711.5023 [M+Na]) (**M13**)	**9** (703.5355 [M+H], 725.5180 [M+Na]) (**M13**)	**15**C/D (747.5617 [M+H]) (**M21**)	**15**A/B (761.5774 [M+H]) (**M21**)	**13**C/D (891.6403 [M+H])	**13**A/B (905.6560 [M+H])		
20200827-57-KS_{14} swap /OE/Δ479-1 (50ml*2)	33248315 (8.9)	90059352 (24)	304965795 (80)	585412754 (153)	n.d.	n.d.	n.d.	121298 (0.21)	8	260
20200827-58-KS_{14} swap /OE/Δ479-2 (50ml*2) ✓	27792311 (7.5)	101651512 (29)	245985991 (65)	615185973 (161)	n.d.	n.d.	n.d.	278506 (0.26)	7.5	260
20200827-64-$KS_{14/21}$ /OE/Δ479-1 (50ml*2)	57473515 (15)	215089188 (57)	694181285 (182)	513896986 (135)	15066416 (4.1)	794478 (0.4)	n.d.	619726 (0.35)	7.5	260
20200827-65-$KS_{14/21}$ /OE/Δ479-2 (50ml*2) ✓	133697567 (35)	649046324 (170)	1295140013 (339)	2231045625 (582)	3694680 (1.2)	276172 (0.26)	n.d.	5188594 (1.5)	8	260

✓ The marked entries refer to the highest yields of metabolites produced by the corresponding mutants.

MS analysis was performed on Thermo Fisher LTQ Orbitrap mass spectrometer prior to 2020, whereas subsequently, the MS analysis and the standardization were carried out on a Thermo Fisher Orbitrap ID-X Tribrid mass spectrometer. As the later mass spectrometer is more sensitive, all of the data acquired with the first instrument were re-scaled by multiplying by 10 (a factor of 10 was determined by analysis of identical samples on the two machines) prior to yield calculation. In addition, as there was substantial biological variability with fermentation of certain strains (with compounds present above or below the detection limits), we chose not to calculate an average yield, but rather to present the individual data.

Supplementary Table 6 List of new metabolites present in various engineered mutants containing the overexpressed regulator, by comparison to the corresponding mutants containing empty plasmid pIB139 and WT/OE484.

Mutants	New metabolites (*m/z*)
K7N1/OE484	No stambomycins 532.3440, 444.2592, 458.2838(+2H), 651.3896, 576.4211, 419.2992, 474.3443(+2H), 425.3361, 480.8504(+2H), 386.7922(+2H), 705.5137, 719.5290, 660.4785
K7N2/OE484	No stambomycins 1348.429, 1371.437, 705.5142, 719.5297, 1364.929, 1370.433, 1518.038
K7N4/OE484	Stambomycins 661.9570 (+2H), 695.4724 (+2H), 697.4833 (+2H), 669.9518 (+2H), 695.4736 (+2H), 699.9836 (+2H), 664.4550 (+2H), 683.9503 (+2H), 658.9611 (+2H), 698.9536 (+2H), 660.4896 (+2H), 672.4711 (+2H), 678.9785 (+2H), 796.5922, 671.4810, 670.9792, 547.4097 (+2H), 556.4149 (+2H)
K7N5/OE484	NO stambomycins 546.8807 (+2H), 553.8887 (+2H), 705.5132, 719.5291
K7N6/OE484	Same with WT/pOE484, no new metabolites
CPN1/OE484	No stambomycins 705.5130, 719.5289, 858.6297
CPN2/OE484	No stambomycins 705.5134, 719.5291, 524.3039 (+2H), 543.4360, 783.5574

All masses are presented in order of retention time, while the masses highlighted in red are consistent with predicted shunt products.

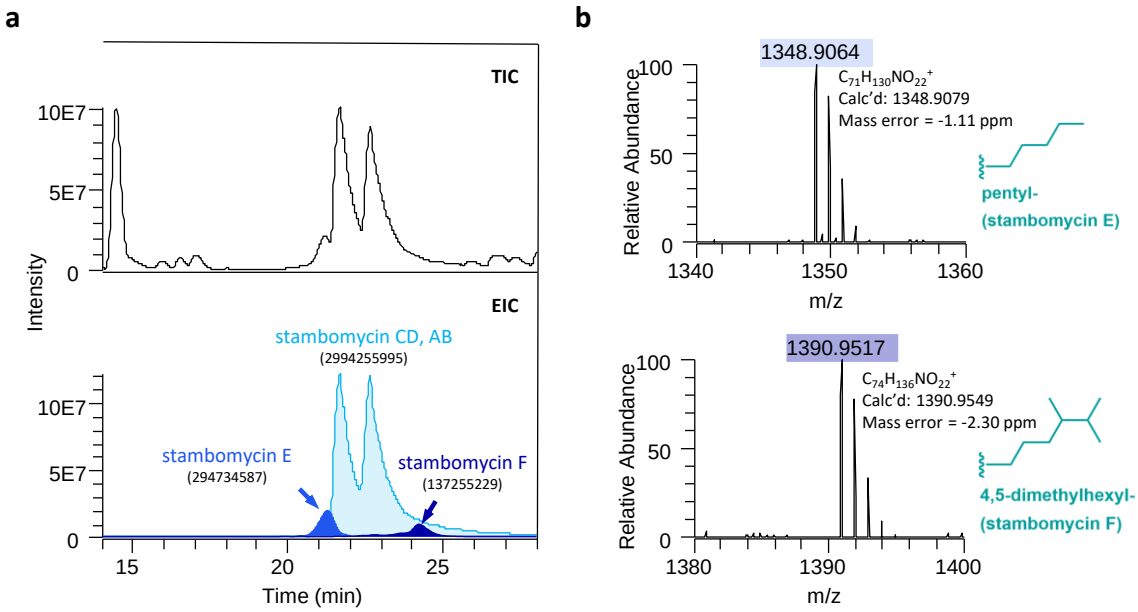
Supplementary References

1. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723−731 (2003).
2. Buchholz, T. J., Geders, T. W., Bartley, F. E., Reynolds, K. A., Smith, J. L. & Sherman, D. H. Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41−52 (2009).
3. Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931−943 (2007).
4. Cobb, R. E., Wang, Y. J. & Zhao, H. M. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**, 723−728 (2015).
5. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4**, 1020−1029 (2015).
6. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207−220 (2001).
7. Böhm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5**, 191−195 (1992).
8. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J.D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16**, 2093−2107 (2007).
9. Kapur, S., *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4110−4115 (2012).
10. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22066−22071 (2010).
11. Wlodek, A*., et al.* Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nat. Comm.* **8**, 1206 (2017).
12. Yuzawa, S*., et al.* Comprehensive *in vitro* analysis of acyltransferase domain exchanges in modular polyketide synthases and its application for short-chain ketone production. *ACS Synth. Biol.* **6**, 139−147 (2017).
13. Song, L., Laureti, L., Corre, C., Leblond, P., Aigle, B. & Challis, G. L. Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis. *J. Antibiot.* **67**, 71−76 (2014).
14. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1541−1546 (2003).
15. Raynal, A., Karray, F., Tuphile, K., Darbon-Rongère, E. & Pernodet, J.-L. Excisable cassettes: new tools for functional analysis of *Streptomyces* genomes. *Appl. Environ. Microbiol.* **72**, 4839−4844 (2006).
16. Zelyas, N., Tahlan, K. & Jensen, S. E. Use of the native flp gene to generate in-frame unmarked mutations in *Streptomyces* spp. *Gene* **443**, 48−54 (2009).
17. Wilkinson, C. J*., et al.* Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417−426 (2002).
18. Laureti, L*., et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6258−6263 (2011).
19. Pinnert-Sindico, S. [A new species of *Streptomyces* producing antibiotics *Streptomyces ambofaciens* n. sp*.,* cultural characteristics]. *Ann. Inst. Pasteur* **87**, 702−707 (1954).
20. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557−580 (1983).
21. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640−6645 (2000).
22. MacNeil, D. J., *et al*. Analysis of *Streptomyces-avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61−68 (1992).

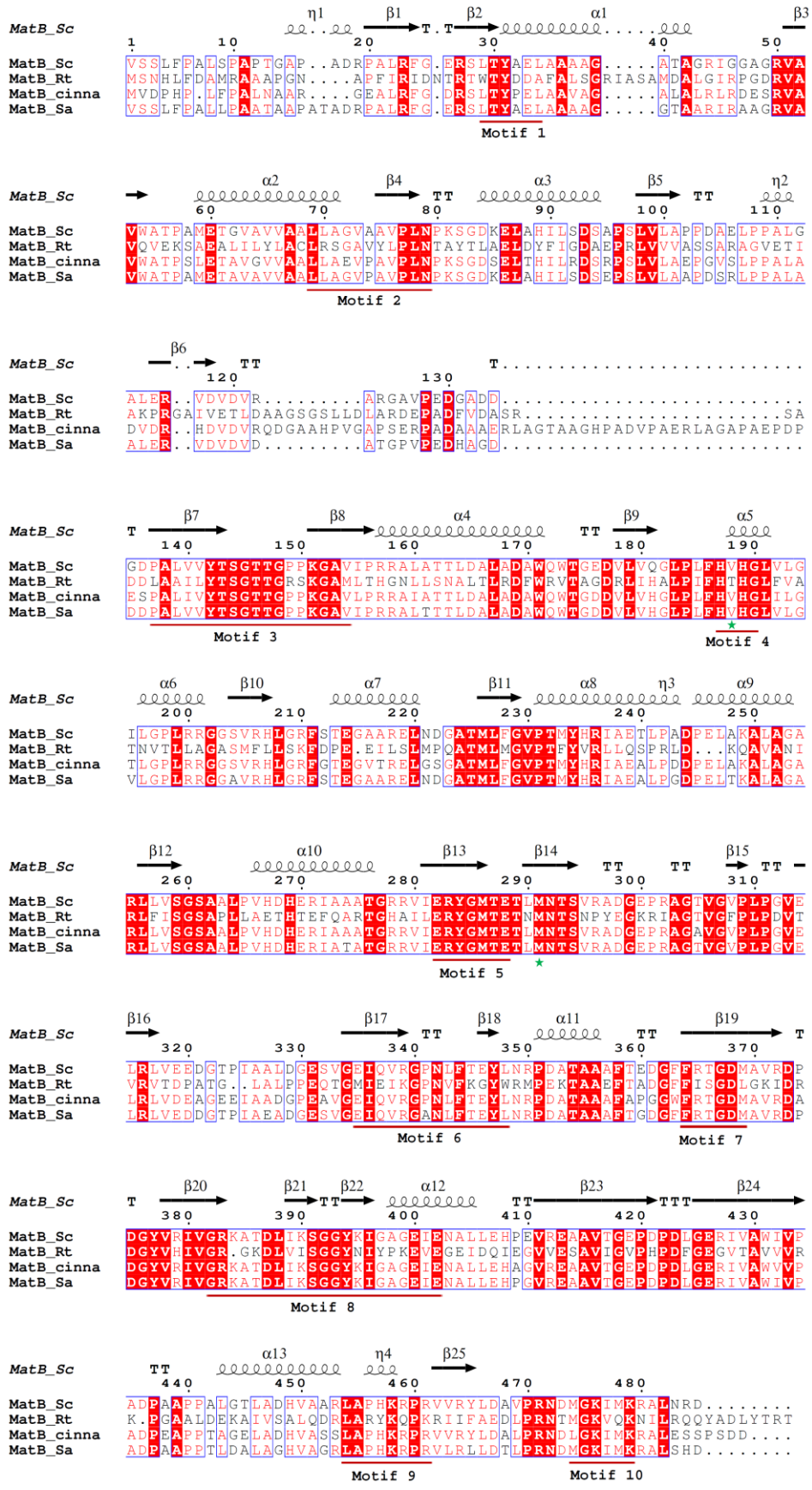
Appendix to Chapter III

Appendix Table 3.1 High resolution LCMS retention times, calculated masses, and observed masses for mutasynthesis of stambomycin analogous. n.d. = not detected

Compound	Retention Time (min)	Calculated Mass $[M+2H]^{2+}$	Observed Mass $[M+2H]^{2+}$ (mass error in ppm)	Calculated Mass $[M+H]^{+}$	Observed Mass $[M+H]^{+}$ (mass error in ppm)
Ethyl-stambomycin	n.d.	653.9345	n.d.	1306.8610	n.d.
Ethyl-deoxystambomycin	19.1	645.9345	n.d.	1290.8660	1290.8647 (–1.00 ppm)
Isopropyl-stambomycin	n.d.	660.9419	n.d.	1320.8766	n.d.
Isopropyl-deoxystambomycin	n.d.	652.9445	n.d.	1304.8817	n.d.
Allyl-stambomycin	19.3	659.9345	n.d.	1318.8610	1318.8606 (–0.30 ppm)
Allyl-deoxystambomycin	19.5	651.9370	651.9361 (–1.38 ppm)	1302.8660	1302.8650 (–0.76 ppm)
Butyl-stambomycin	20	667.9502	667.9493 (–1.35 ppm)	1334.8923	1334.8885 (–2.84 ppm)
Butyl-deoxystambomycin	20.4	659.9527	659.9521 (–0.91 ppm)	1318.8973	1318.8948 (–1.89 ppm)
Benzyl-stambomycin	21.5	684.9580	684.9417 (–23.8 ppm)	1368.8766	1368.8782 (1.17 ppm)
Benzyl-deoxystambomycin	21.5	676.9423	676.9438 (2.22 ppm)	1352.8817	1352.8781 (–2.66 ppm)
Octyl-stambomycin	25.4	695.9811	695.9805 (–0.86 ppm)	1390.9547	1390.9546 (–0.07 ppm)
Octyl-deoxystambomycin	n.d.	687.9836	n.d.	1374.9599	n.d.
Phenylpropyl-stambomycin	n.d.	698.9576	n.d.	1396.9079	n.d.
Phenylpropyl-deoxystambomycin	n.d.	690.9602	n.d.	1380.9130	n.d.
Phenoxypropyl-stambomycin	22.5	706.9554	n.d.	1412.9028	1412.8993 (–2.48 ppm)
Phenoxypropyl-deoxystambomycin	22.5	698.9555	n.d.	1396.9079	1396.9048 (–2.22 ppm)
Thiophene-stambomycin	n.d.	680.9123	n.d.	1360.8174	n.d.
Thiophene-deoxystambomycin	n.d.	672.9148	n.d.	1344.8225	n.d.
Thienylmethyl-stambomycin	n.d.	687.9201	n.d.	1374.8330	n.d.
Thienylmethyl-deoxystambomycin	n.d.	679.9227	n.d.	1358.8381	n.d.
Propargyl-stambomycin	n.d.	658.9263	n.d.	1316.8453	n.d.
Propargyl-deoxystambomycin	n.d.	650.9289	n.d.	1300.8504	n.d.
6-Bromohexyl-stambomycin	22.6	720.9207	720.9130 (–10.7 ppm)	1440.8341	1440.8217 (–8.6 ppm)
6-Bromohexyl-deoxystambomycin	n.d.	712.9232	n.d.	1424.8391	n.d.



Appendix Figure 3.1 Analysis by mass spectrometry of crude extract from wild type ATCC/OE484. a) TIC of the crude extract of ATCC/OE484 and the EIC of stambomycin A–F. The production of stambomycin E and F are only 10% and 5% relative to stambomycins A–D, as judged on the basis of ion intensity. **b**) Mass spectra of stambomycins E and F.


MatB_Sc
MatB_Rt
MatB_cinna
MatB_Sa
Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 6
Motif 7
Motif 8
Motif 9
Motif 10

Appendix Figure 3.2 Sequence alignment of MatB-type enzymes. The secondary structure of MatB_Sc (PDB 3NYQ) is shown at the top. Residues which are strictly conserved are indicated in white on a red background, while relatively conserved residues are indicated in red. The red underlines represent reserved core motifs for adenylate-forming enzymes, in which, the motifs 4 and 5 are identified to bind the carboxylate substrates and stabilize the formation of the adenylate intermediates [1]. The green stars refer to two amino acids mutated in MatB_Rt, which are hydrophobic residues likely to influence the promiscuity of MatB-type enzymes for α-substituted malonate derivatives on the basis of the elucidated structure of MatB_Sc [2]. MatB_Sa: MatB from S. ambofaciens; MatB_Sc: MatB from *S. coelicolor*; MatB_cinna: MatB from *S. cinnamonensis*; MatB_Rt: MatB from *R. trifolii*. MatB_Sa exhibits 90%, 75% and 39% amino acid sequence identity to MatB_Sc, MatB_cinna and MatB_Rt, respectively.

Appendix References to chapter III

1. Hughes, Amanda J. and A. Keatinge-Clay, *Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Func-tional Showcasing of Streptomyces coelicolor MatB.* Chemistry & Biology, 2011. **18**(2): p. 165-176.
2. Koryakina, I., et al., *Poly Specific trans-Acyltransferase Machinery Revealed via Engineered Acyl-CoA Synthetases.* ACS Chemical Biology, 2013. **8**(1): p. 200-208.

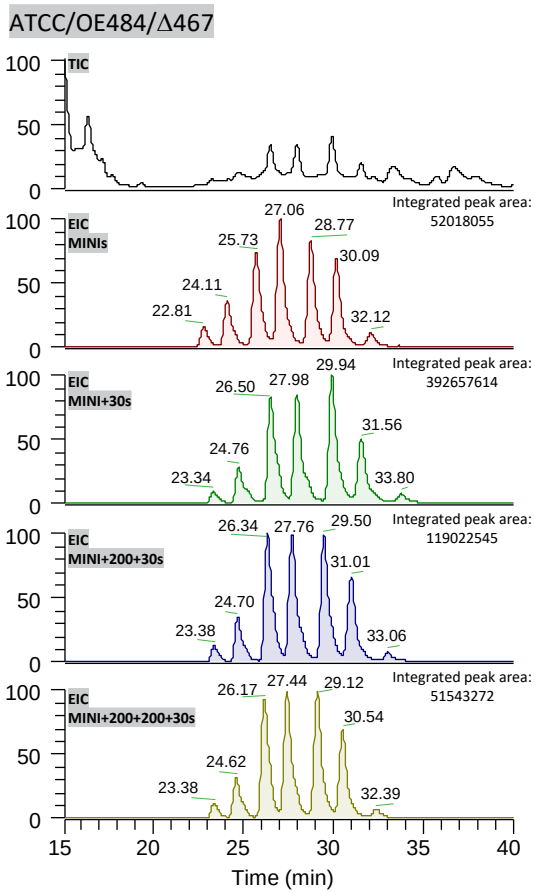
Appendix to Chapter IV

Appendix Table 4.1 Enzymes encoded by the *des* cluster for DFO biosynthesis in *S. ambofaciens*, compared to their homologs in *S. coelicolor*.

Name	Features	AA identities (relative to *S. coelicolor*)	Accession number
DesA_Sa	putative L-Lysine decarboxylase	95%	AKZ55866.1
DesB_Sa	FAD-dependent amine *N*-monooxygenase	91%	AKZ55867.1
DesC_Sa	putative acyl-CoA acyltransferase	83%	AKZ55868.1
DesD_Sa	putative Type C siderophore synthetase	91%	AKZ55869.1
DesE_Sa	cell surface-associated lipoprotein receptor	92%	AKZ55864.1
DesF_Sa	putative ferric-siderophore hydrolase	87%	AKZ55865.1

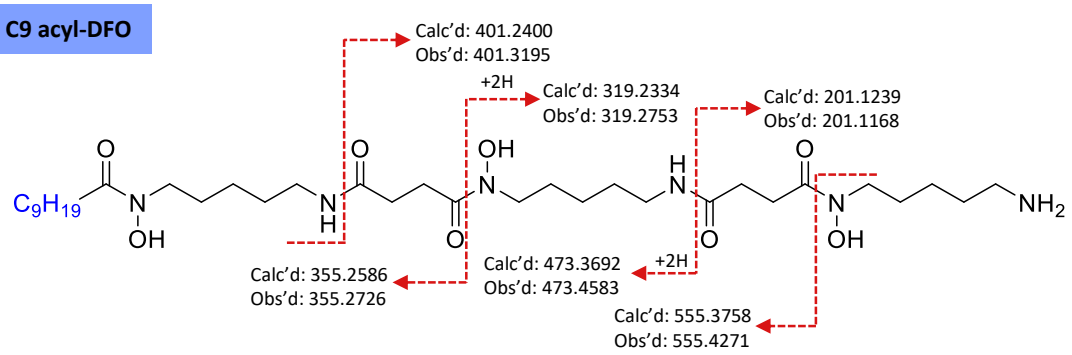
Appendix Table 4.2 Comparison of the yield of DFOs in different strains based on the integrated EIC peak areas.

Strains (culture volume)	Integrated area of MINIs (673.4858, 687.5015, 701.5171, 715.5328, 729.5481, 743.5641, 757.5797)	Integrated area of MINIs+30 (703.4600, 717.4757, 731.4913, 745.5070, 759.5226, 773.5383, 787.5539)	Integrated area of MINIs+200+30 (903.5758, 917.5918, 931.6074, 945.6231, 959.6387, 973.6544, 987.6700)	Integrated area of MINIs+200 +200+30 (1103.6919, 1117.7079, 1131.7235, 1145.7392, 1159.7548, 1173.7705, 1187.7861)	Mass of cells (g)	Volume of final extract in MeOH (µL)
20190130-9-WTOE (50 ml)	641338916	886696662	77873602	11011498	3.8	300
20190206-11-WTOE (50 ml)	565101284	683678314	74281989	11689054	3.5	250
Average aera	603220100	785187488	76077795	11350276		
20190206-9-K7N1/OE (50 ml)	514511930	1058664012	289131723	102161531	3	210
20190405-22-K7N1/OE (50 ml)	862871842	1268029489	449262473	212626507	2.2	170
Average aera	688691886	1163346750	369197098	157394019		
20190405-26-CPN1/OE (50 ml)	689040238	1501040357	295564572	62717614	3	210
20190529-8-CPN1/OE (50 ml)	850151941	1060385742	226734172	58644231	3.5	250
Average aera	769596090	1280713049	261149372	60680923		
20180604-ATCC/OE/Δ483-1 (50 ml)	370629948	371455559	116806527	41405952	3.3	210
20180604-ATCC/OE/Δ483-2 (50 ml)	326055274	291581002	79582457	29572825	2.9	200
Average aera	348342611	331518281	98194492	35489389		
20190318-ATCC/OE/Δ467-a (50 ml)	59752259	458627989	136556515	54992120	4.1	308
20190318-ATCC/OE/Δ467-b (50 ml)	52711050	447958420	131865949	51139850	3.8	285
20190318-ATCC/OE/Δ467-c (50 ml)	43590855	271386434	88645170	48497847	4.5	338
Average aera	52018055	392657614	119022545	51543272		

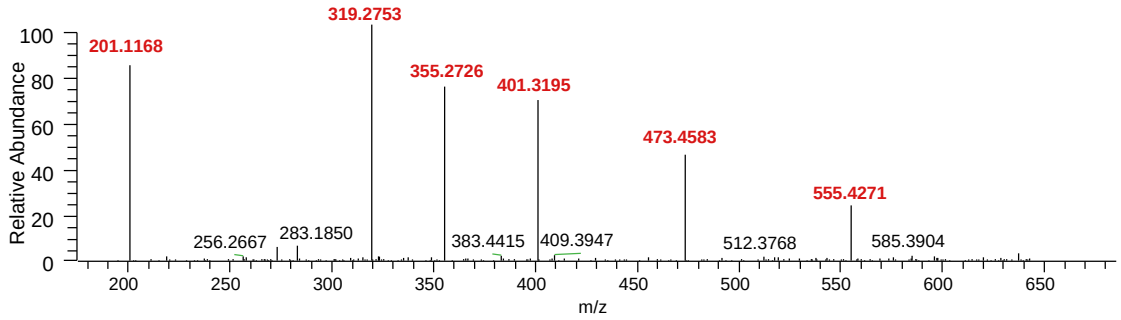


Appendix Figure 4.1 HPLC-MS analysis of acyl-DFOs and their various derivatives (acyl-DFOs+30, acyl-DFOs+200+30, and acyl-DFOs+200+200+30) from mutant ATCC/OE484/Δ467. TIC (total ion chromatogram) of the mutant extracts and the EIC (extracted ion chromatogram) of acyl-DFOs, acyl-DFOs+30, acyl-DFOs+200+30, and acyl-DFOs+200+200+30 are shown. Based on the integrated EIC peak area, the yield of all DFO derivatives in this strain was estimated to be 26% of that produced by the DD engineering mutants.

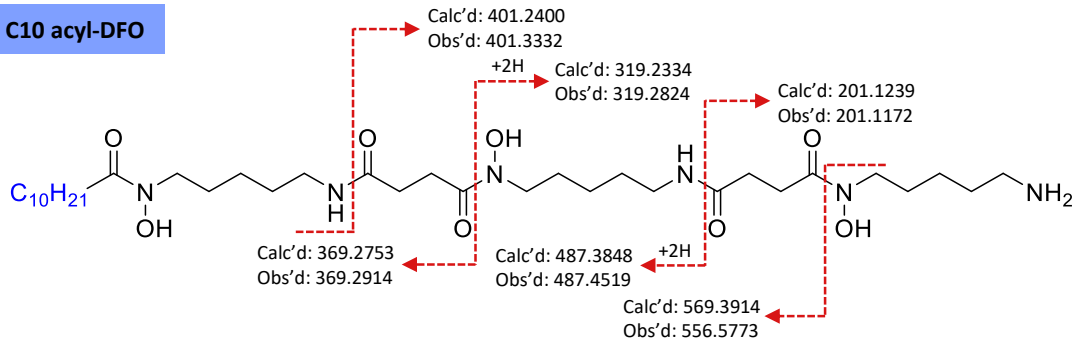
C9 acyl-DFO



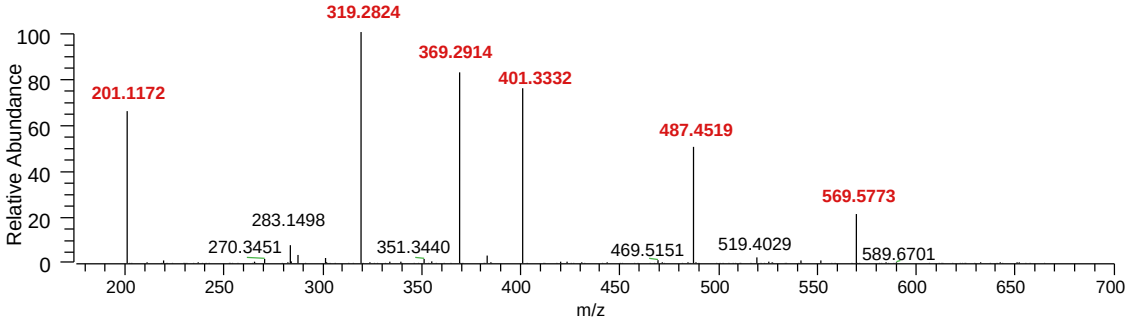
dynamic-ls-20190405-K7N1-oe #1987-2078 RT: 23.25-23.51 AV: 2 NL: 7.18E2
F: ITMS + c ESI d w Full ms2 673.30@cid35.00 [175.00-685.00]



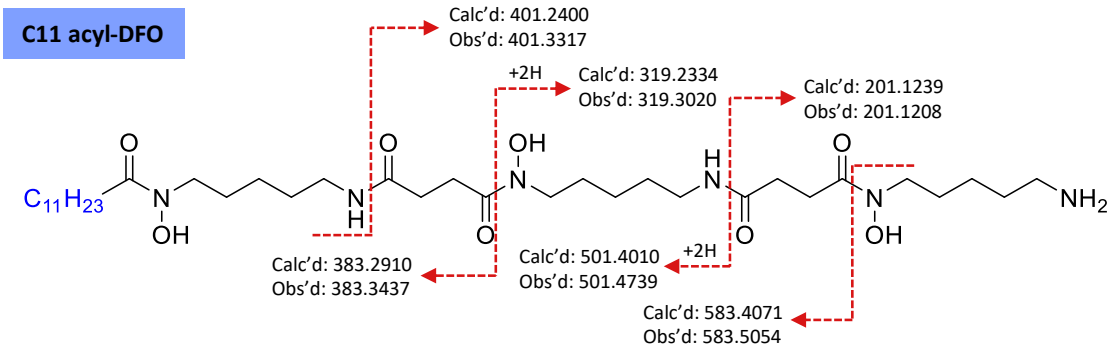
C10 acyl-DFO



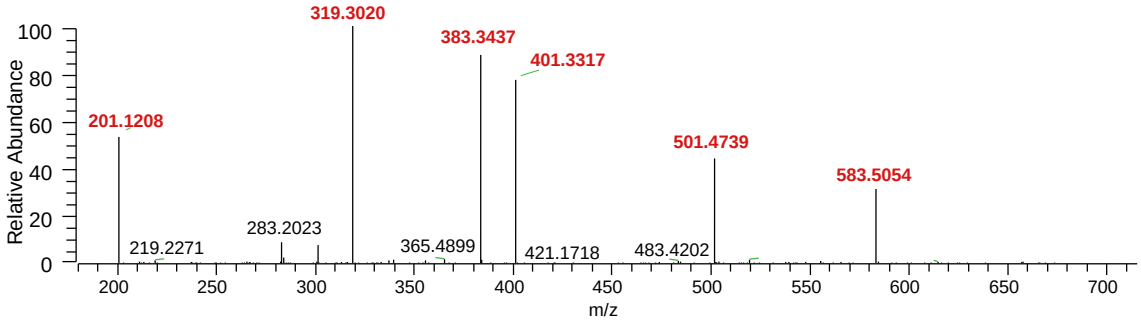
dynamic-ls-20190405-K7N1-oe #2092-2258 RT: 24.30-26.05 AV: 2 NL: 3.40E3
F: ITMS + c ESI d w Full ms2 687.50@cid35.00 [175.00-700.00]



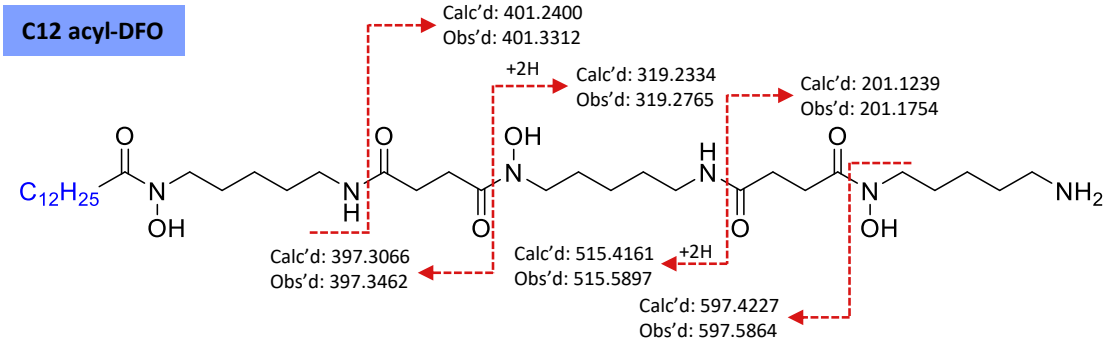
C11 acyl-DFO



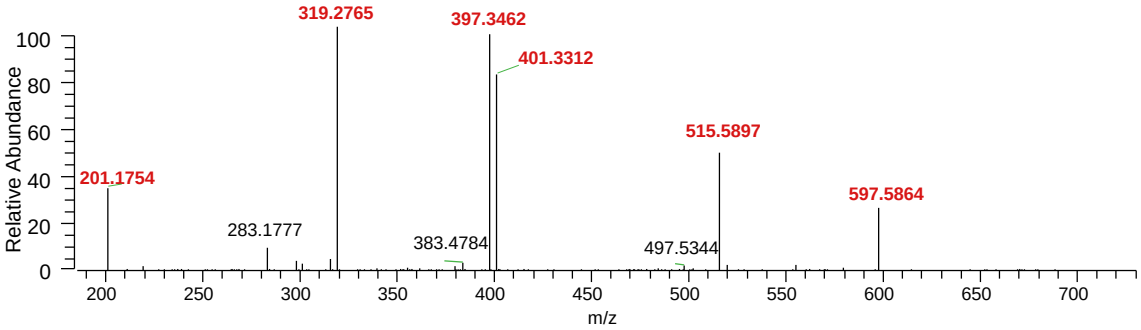
dynamic-ls-20190405-K7N1-oe #2179-2324 RT: 25.89-26.68 AV: 2 NL: 3.97E3
F: ITMS + c ESI d w Full ms2 701.43@cid35.00 [180.00-715.00]



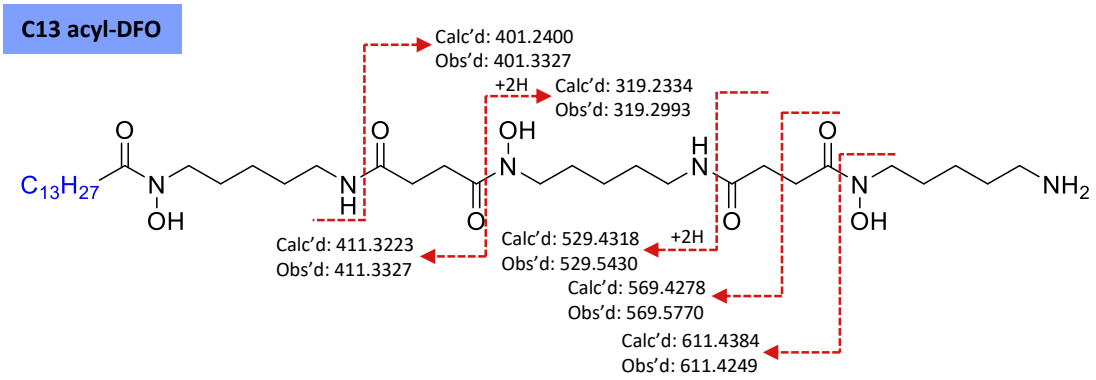
C12 acyl-DFO



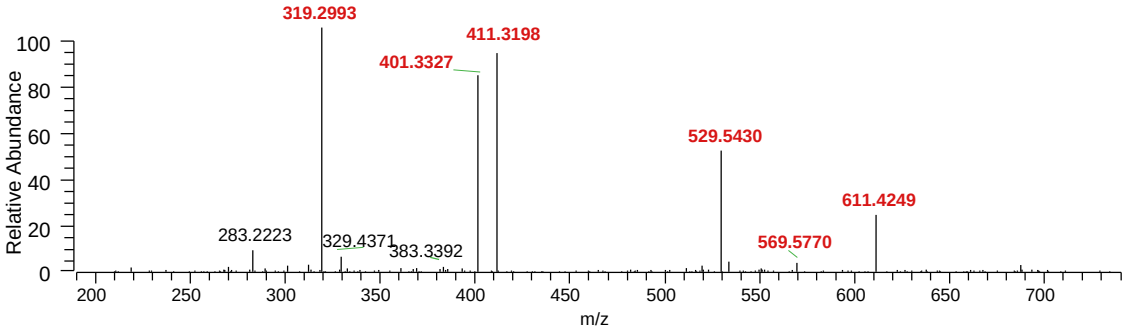
dynamic-ls-20190405-K7N1-oe #2344-2456 RT: 27.20-28.01 AV: 2 NL: 1.56E4
F: ITMS + c ESI d w Full ms2 715.53@cid35.00 [185.00-730.00]



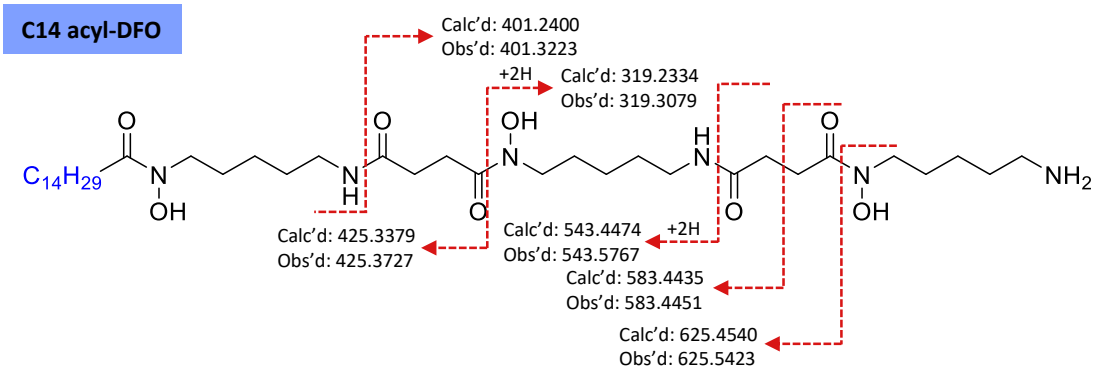
C13 acyl-DFO



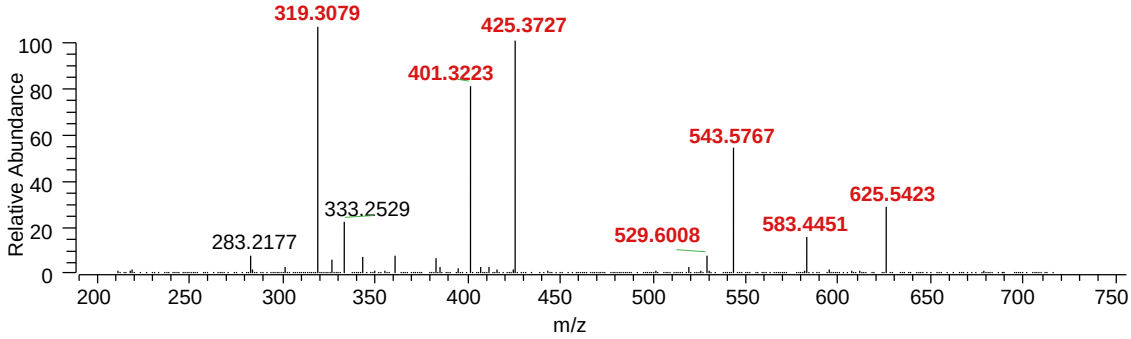
DYNAMIC-LS-20190405-K7N1-oe #2533-2672 RT: 29.12-30.45 AV: 2 NL: 1.38E3
F: ITMS + c ESI d w Full ms2 729.42@cid35.00 [190.00-740.00]

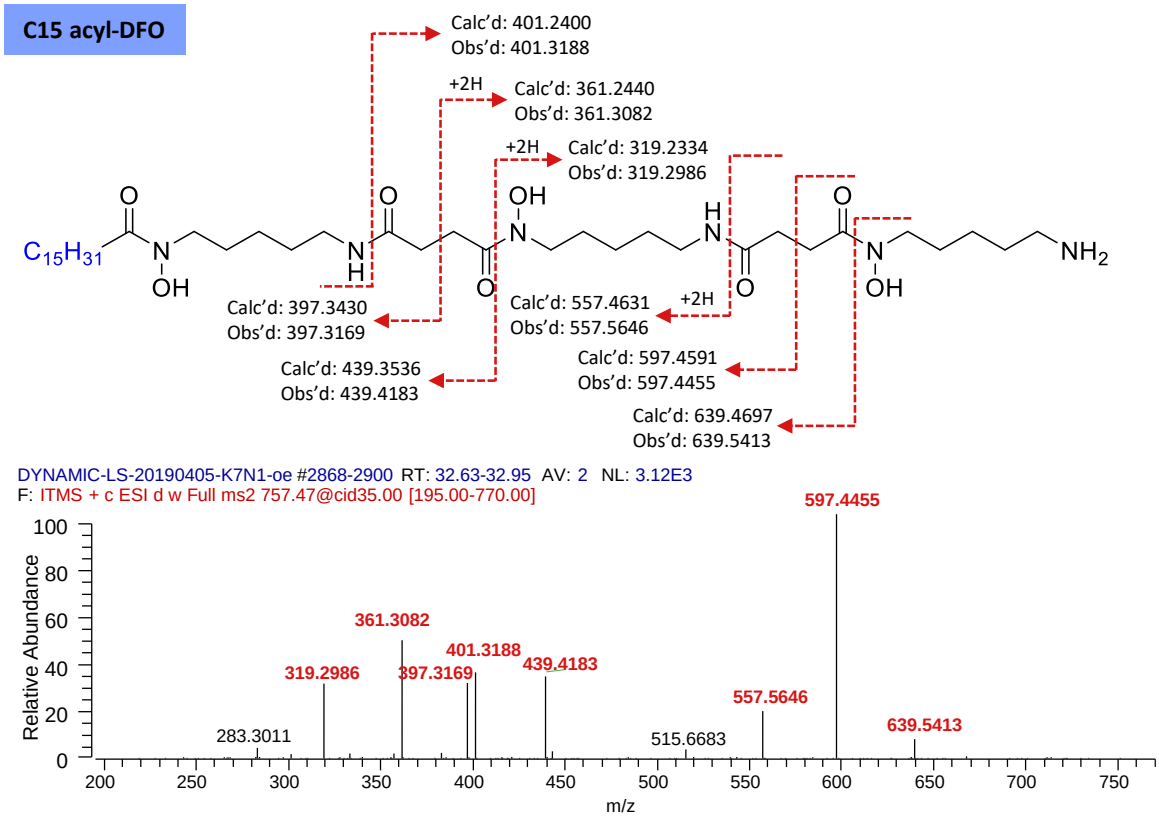


C14 acyl-DFO



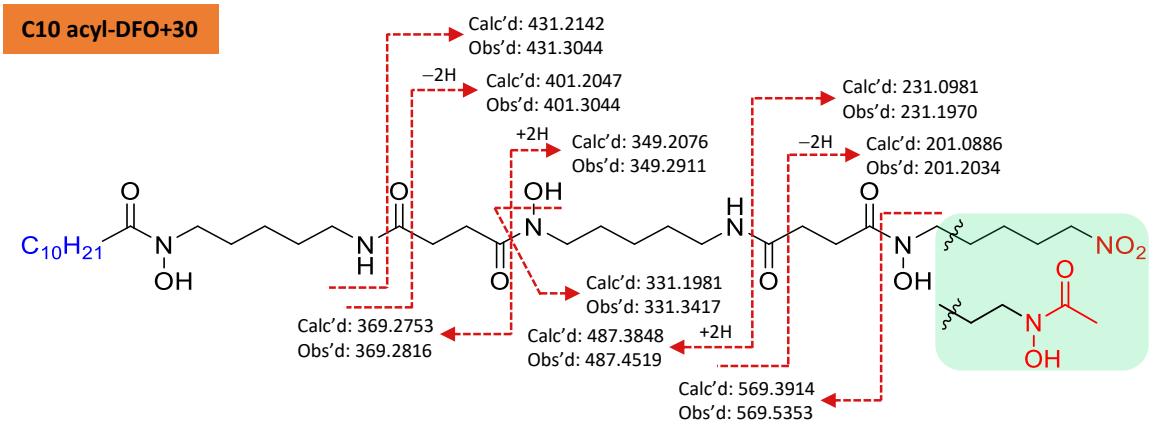
DYNAMIC-LS-20190405-K7N1-oe #2586-2721 RT: 29.76-31.05 AV: 4 NL: 1.79E3
F: ITMS + c ESI d w Full ms2 743.53@cid35.00 [190.00-755.00]



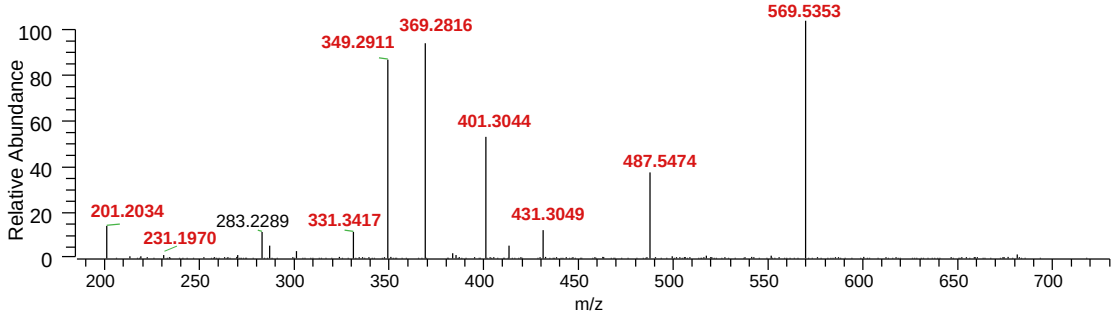


Appendix Figure 4.2 Annotation of fragment spectra of all acyl-DFOs detected in *S. ambofaciens* wild type and mutants. The structures of C9–C15 acyl-DFOs were elucidated by Traxler *et al*. [1].

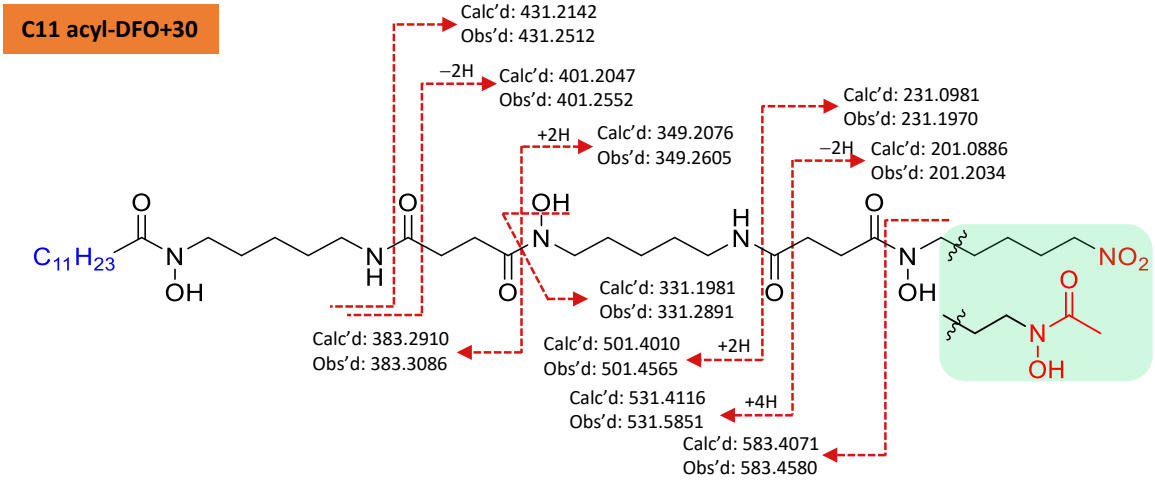
C10 acyl-DFO+30



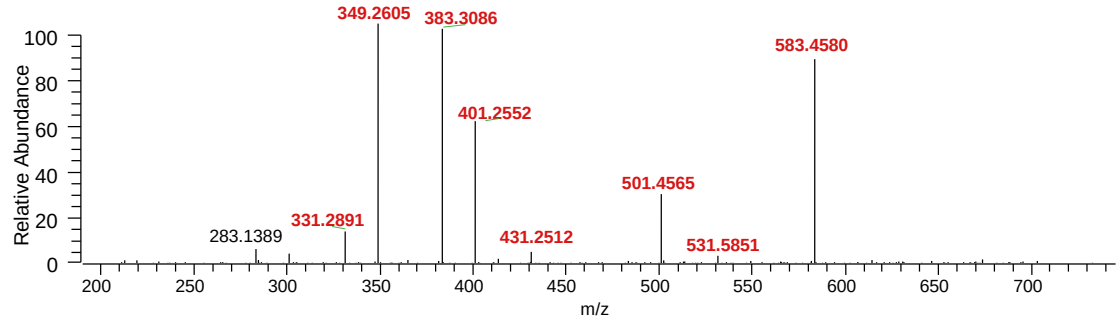
dynamic-ls-20190405-K7N1-oe #2125-2280 RT: 24.67-25.47 AV: 2 NL: 1.64E3
F: ITMS + c ESI d w Full ms2 717.47@cid35.00 [185.00-730.00]



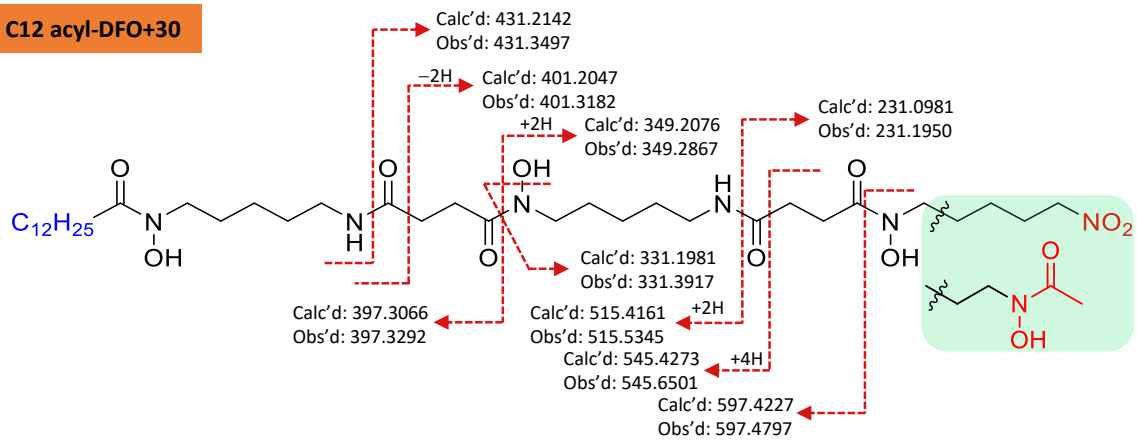
C11 acyl-DFO+30



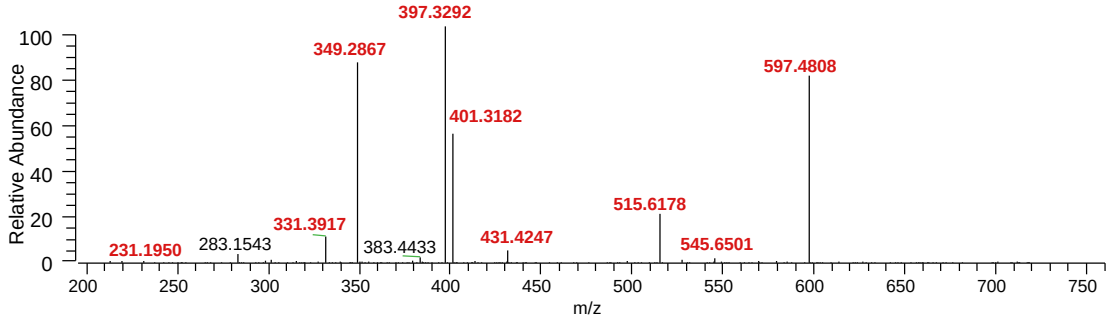
dynamic-ls-20190405-K7N1-oe #2372-2245 RT: 26.38-27.18 AV: 2 NL: 1.26E3
F: ITMS + c ESI d w Full ms2 731.49@cid35.00 [190.00-745.00]



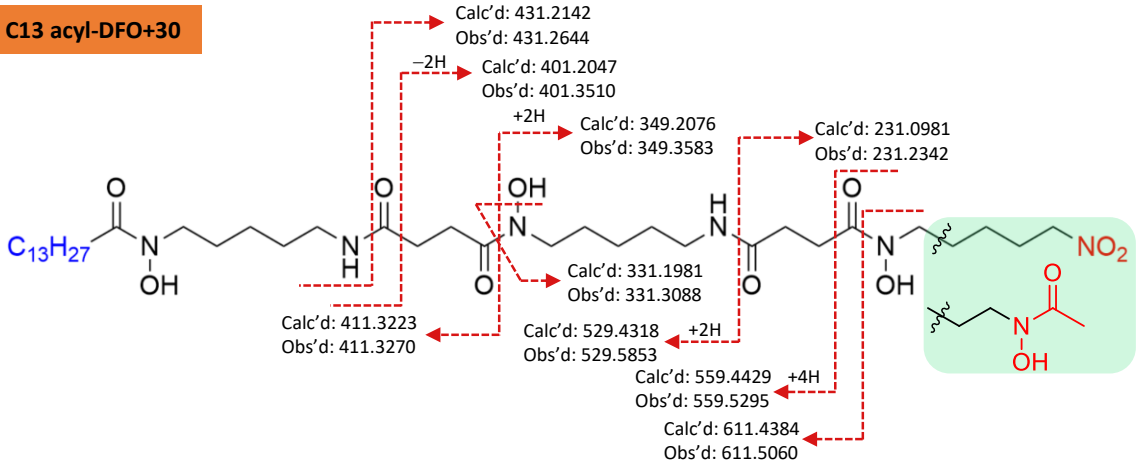
C12 acyl-DFO+30



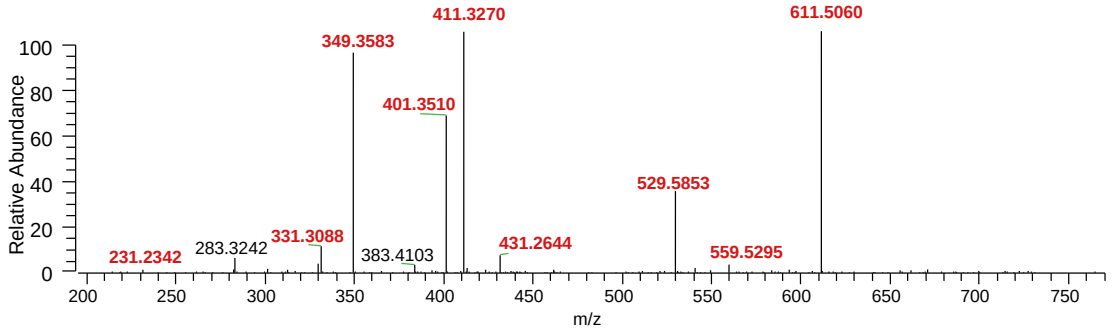
DYNAMIC-LS-20190405-K7N1-oe #2396-2520 RT: 27.85-28.90 AV: 2 NL: 4.93E3
F: ITMS + c ESI d w Full ms2 745.51@cid35.00 [195.00-760.00]

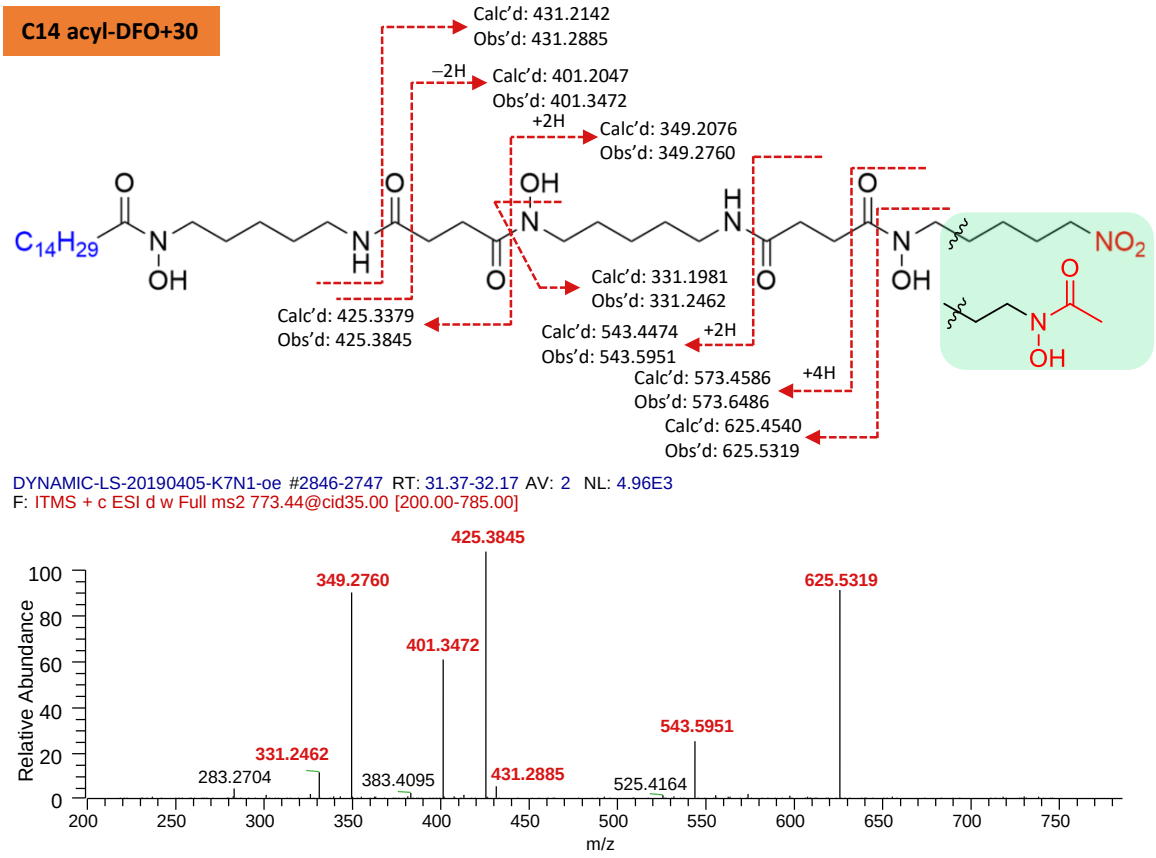


C13 acyl-DFO+30

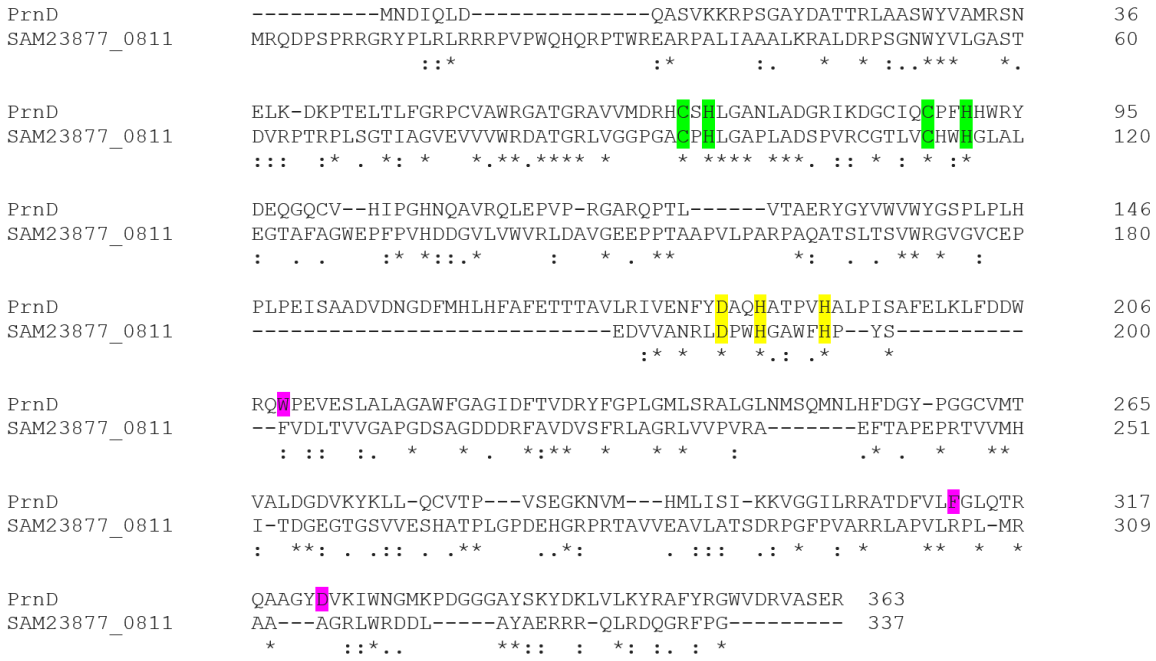


DYNAMIC-LS-20190405-K7N1-oe #2595-2688 RT: 29.79-30.58 AV: 2 NL: 1.23E3
F: ITMS + c ESI d w Full ms2 759.52@cid35.00 [195.00-770.00]



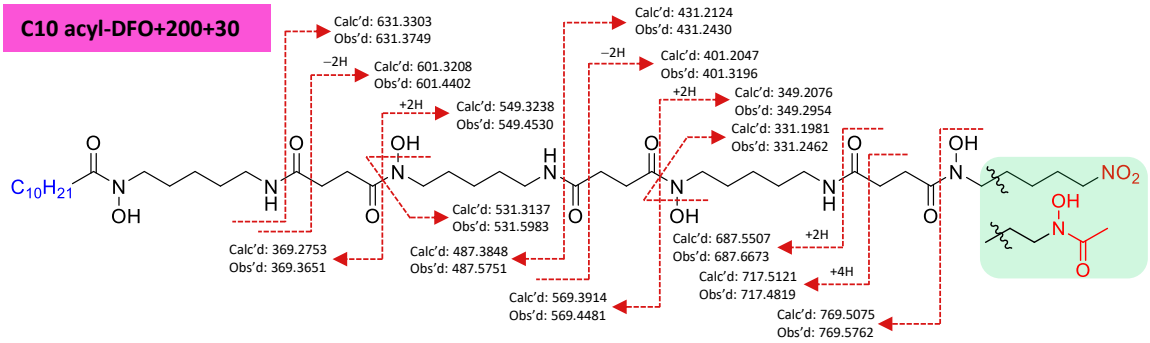


Appendix Figure 4.3 Annotation of fragment spectra of all acyl-DFOs+30 detected in *S. ambofaciens* wild type and mutants. The structures of C10–C14 acyl-DFOs+30 are predicted based on comparison of fragments with acyl-DFOs. The MS/MS spectra of C9 and C15 acyl-DFO+30 are absent, as the intensity of the parental ions was insufficient to trigger fragmentation.

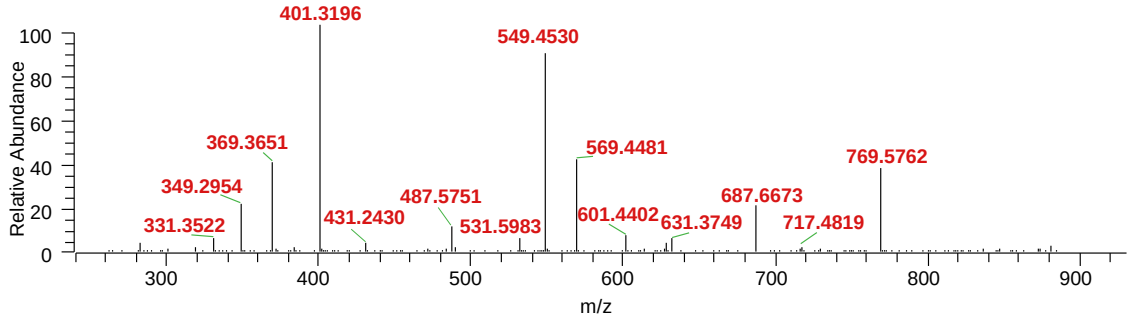


Appendix Figure 4.4 Sequence alignment of PrnD and SAM23877_0811. PrnD is an aminopyrrolnitrin oxygenase involved in the biosynthesis of the antifungal antibiotic pyrrolnitrin. The residue numbers of the two proteins are indicated at the end of each line. The DxxHxxxxH (highlighted in yellow) is the conserved motif for [2Fe-2S] Rieske non-heme Fe-protein an involved in binding oxygen in oxygenases. Residues 69C, 71H, 88C and 91H (highlighted in green) of PrnD are the metal binding sites, and 209W, 312F and 323D are residues identified to be significant for oxygenase activity using mutagenesis [2, 3].

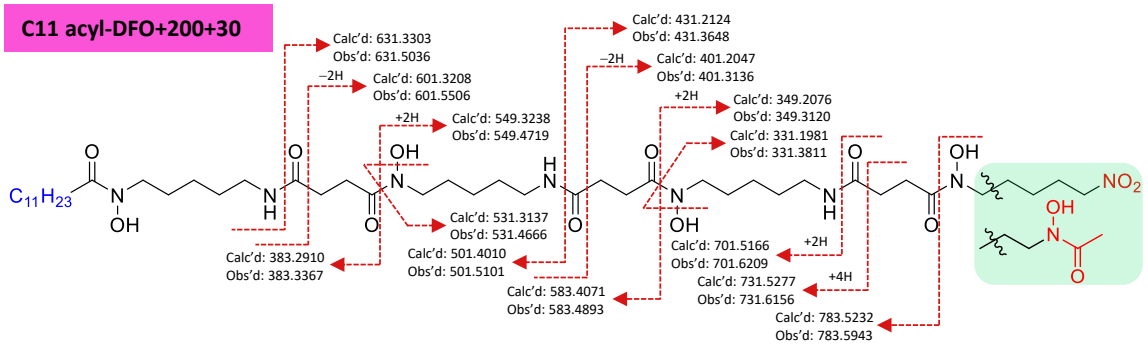
C10 acyl-DFO+200+30



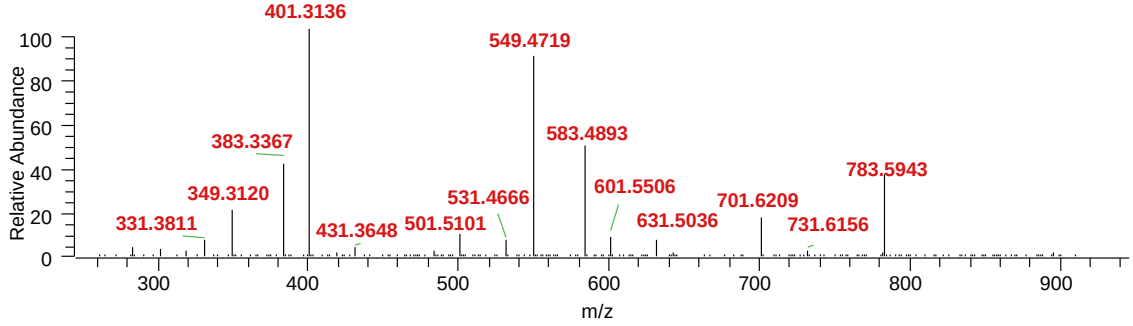
DYNAMIC-LS-20190405-K7N1-oe #2130 RT: 24.69 AV: 1 NL: 1.26E3
F: ITMS + c ESI d w Full ms2 917.49@cid35.00 [240.00-930.00]



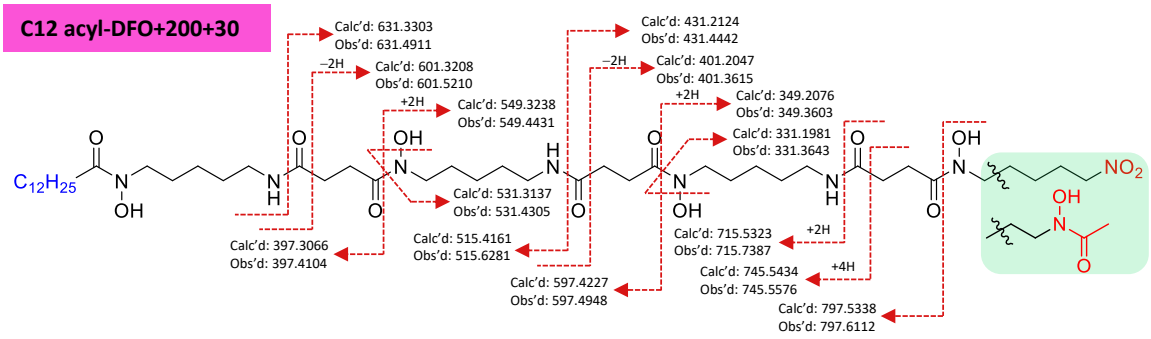
C11 acyl-DFO+200+30



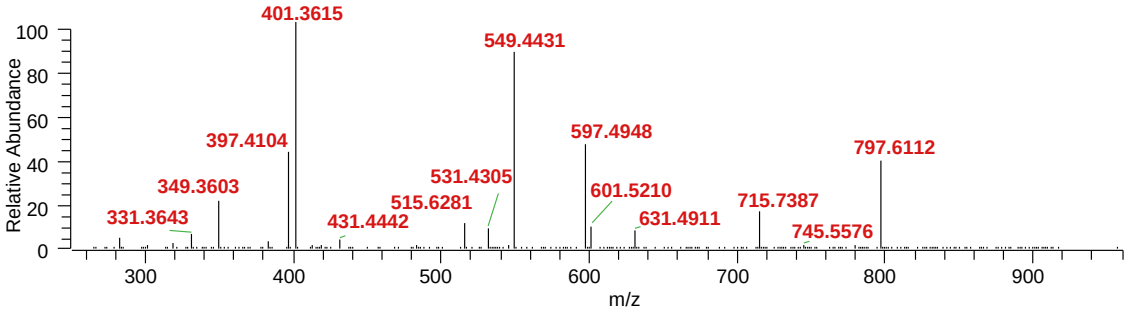
DYNAMIC-LS-20190405-K7N1-oe #2262-2307 RT: 26.27-26.53 AV: 2 NL: 2.10E3
F: ITMS + c ESI d w Full ms2 931.61@cid35.00 [245.00-945.00]



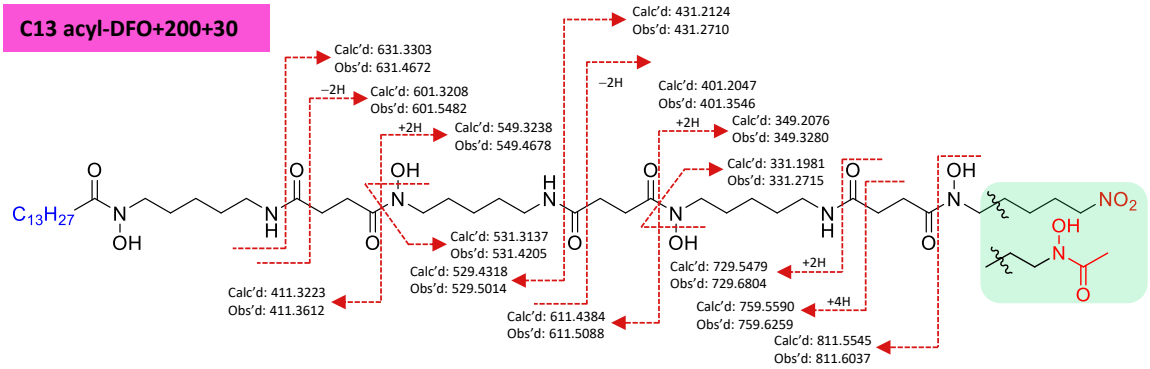
C12 acyl-DFO+200+30



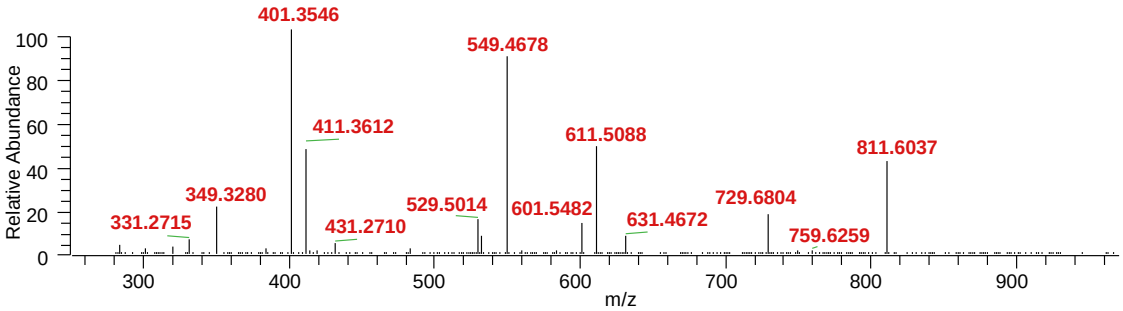
DYNAMIC-LS-20190405-K7N1-oe #2389-2467 RT: 27.87-28.14 AV: 2 NL: 3.60E3
F: ITMS + c ESI d w Full ms2 945.62@cid35.00 [250.00-960.00]

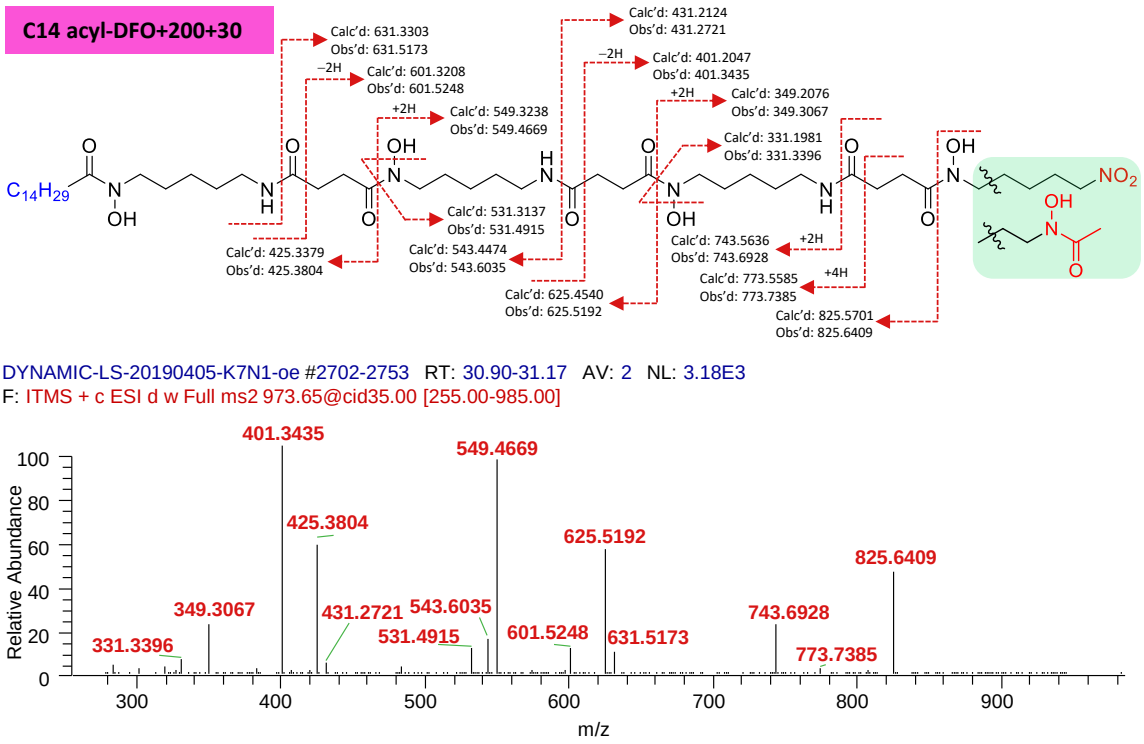


C13 acyl-DFO+200+30



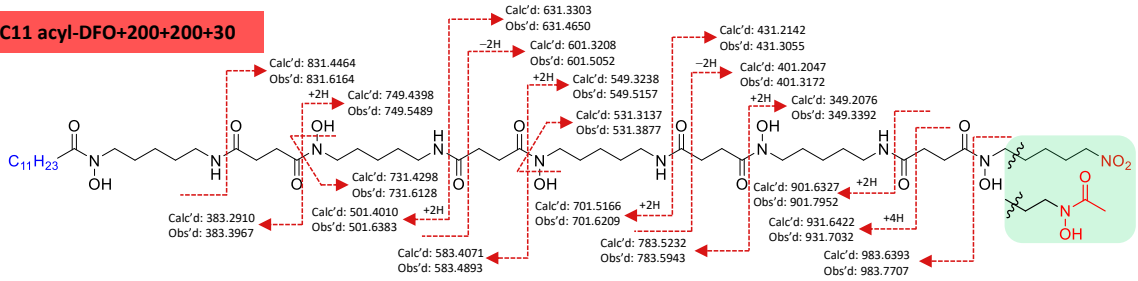
DYNAMIC-LS-20190405-K7N1-oe #2564-2608 RT: 29.44-29.70 AV: 2 NL: 2.42E3
F: ITMS + c ESI d w Full ms2 959.64@cid35.00 [250.00-970.00]



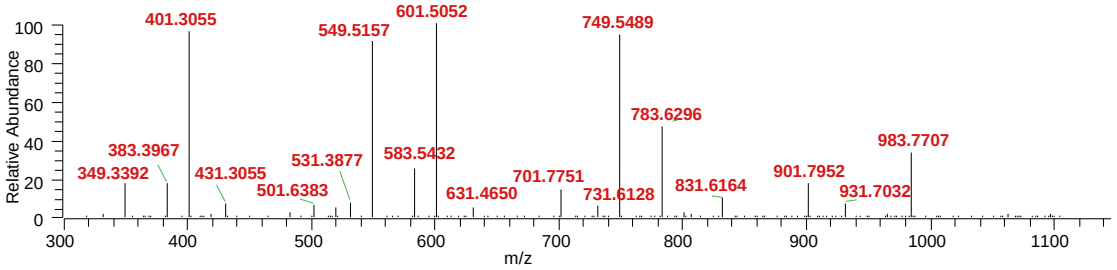


Appendix Figure 4.5 Annotation of fragment spectra of all acyl-DFOs+200+30 detected in *S. ambofaciens* wild type and mutants. The structures of C10–C14 acyl-DFOs+200+30 are predicted based on comparison of fragments with acyl-DFO+30s. The MS/MS spectra of C9 and C15 acyl-DFO+200+30 are absent due to due to their parent masses inferior to the intensity threshold for triggering fragmentation.

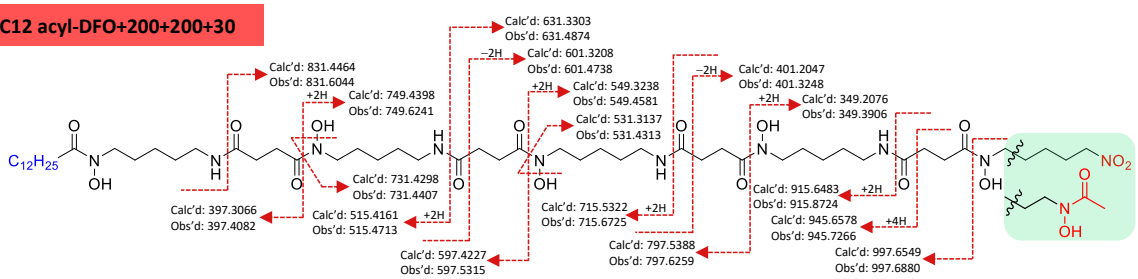
C11 acyl-DFO+200+200+30



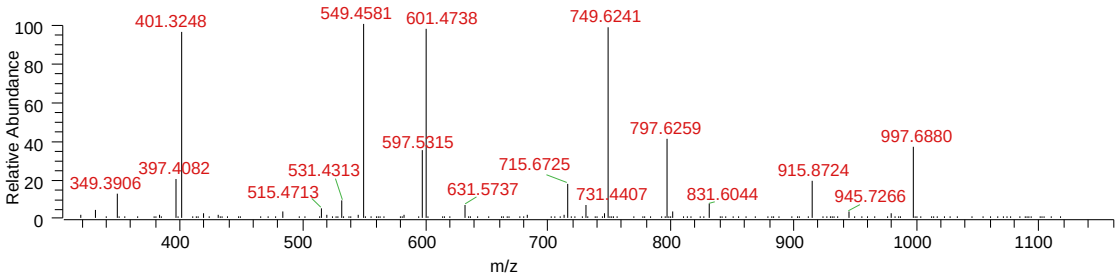
DYNAMIC-LS-20190405-K7N1-oe #2239-2277 RT: 26.11-26.41 AV: 2 NL: 4.85E2
F: ITMS + c ESI d w Full ms2 1131.61@cid35.00 [300.00-1145.00]

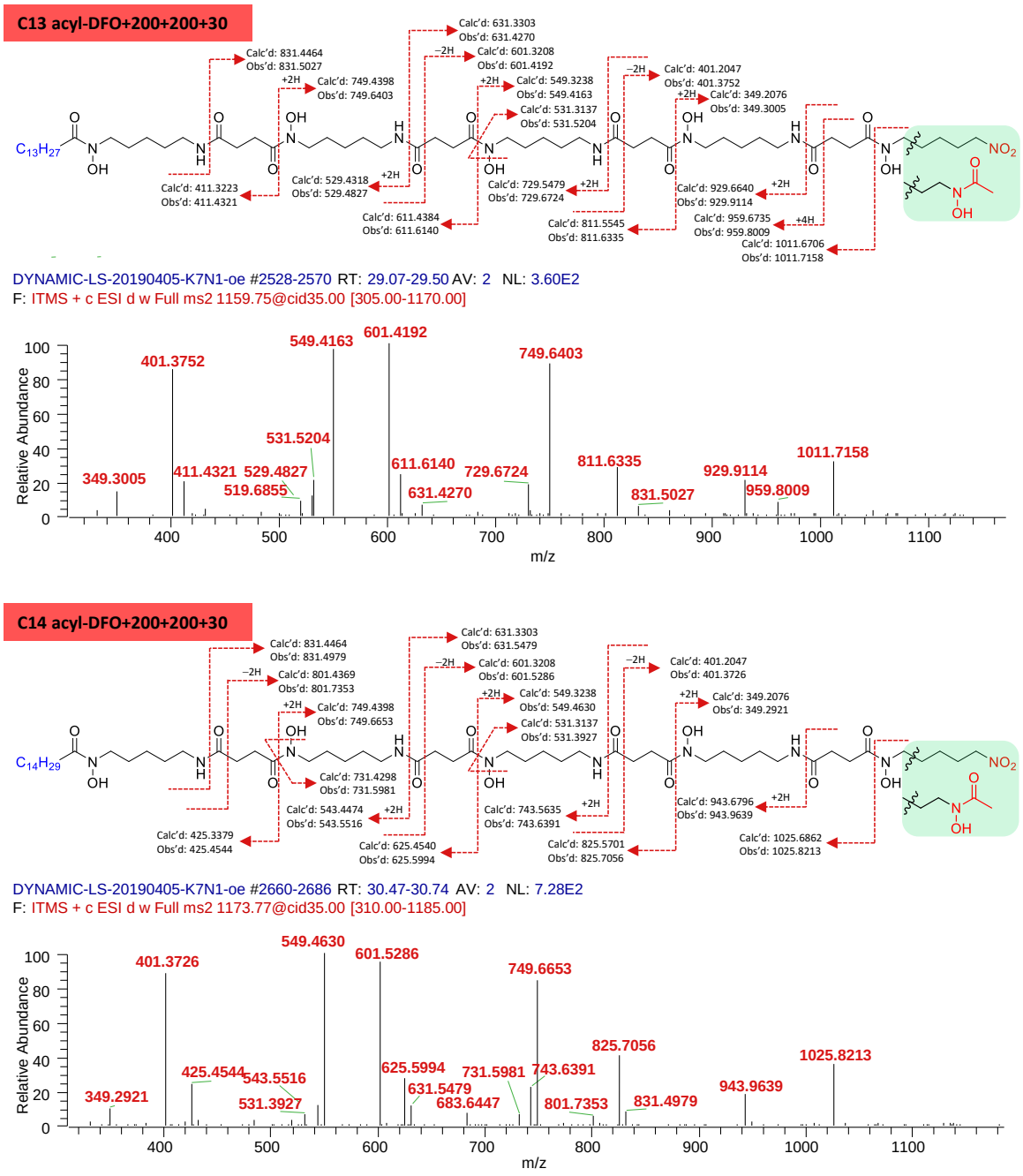


C12 acyl-DFO+200+200+30

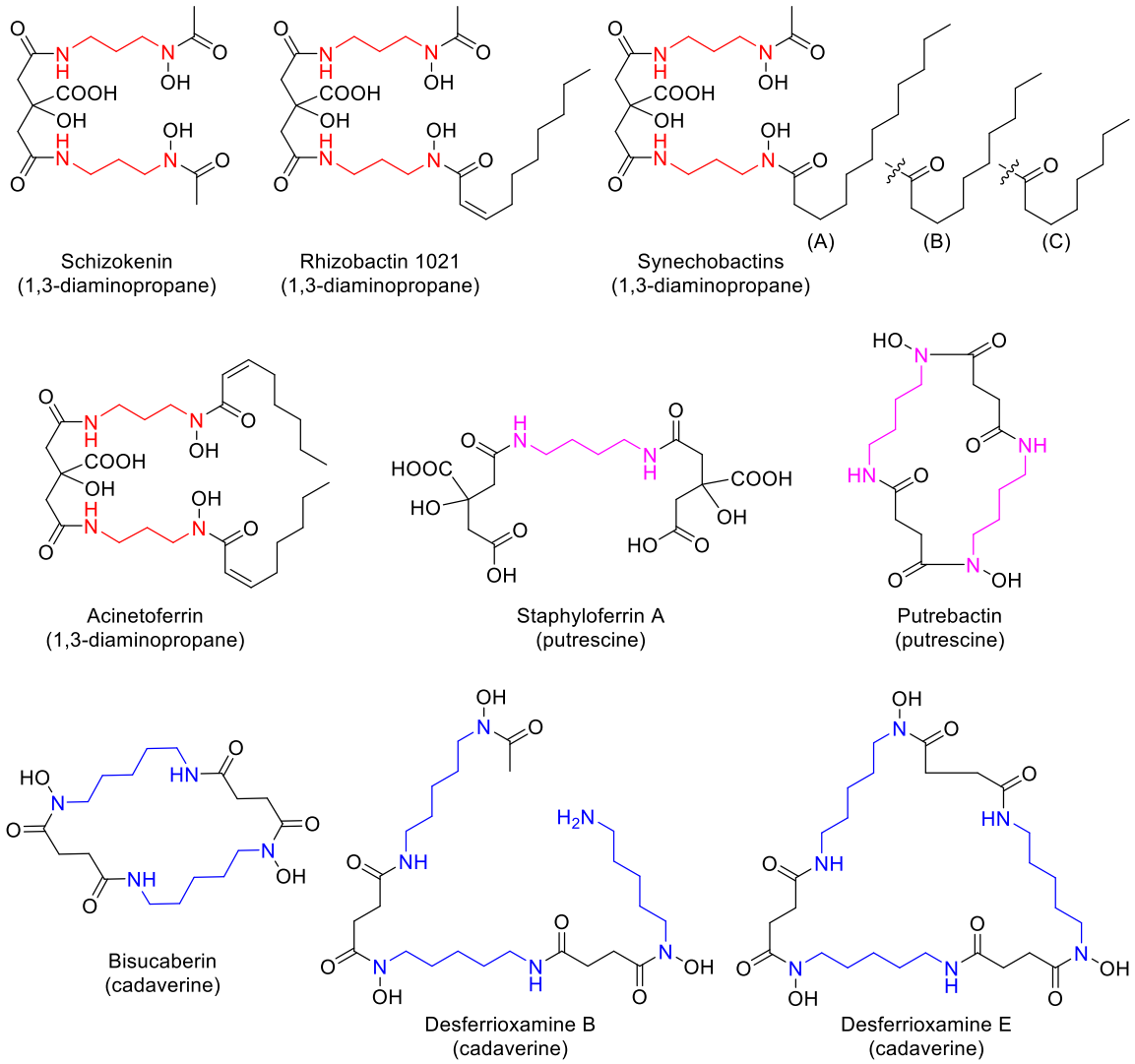


DYNAMIC-LS-20190405-K7N1-oe #2372 RT: 27.45 AV: 1 NL: 1.47E3
F: ITMS + c ESI d w Full ms2 1145.74@cid35.00 [305.00-1160.00]





Appendix Figure 4.6 Annotation of fragment spectra of all acyl-DFOs+200+200+30 detected in *S. ambofaciens* wild type and mutants. The structures of C11–C14 acyl-DFOs+200+30 are predicted based on comparison of fragments with acyl-DFO+30s. The MS/MS spectra of C9, C10 and C15 acyl-DFO+200+200+30 are absent due to due to their parent masses inferior to the intensity threshold for triggering fragmentation.



Appendix Figure 4.7 Structures of NIS-mediated siderophores. The 1,3-diaminepropane backbone is indicated in red, 1,4-diaminepropane (putrescine) in pink, 1,5-diaminepropane (cadaverine) in blue, with the backbones indicated underneath the siderophore names.

Appendix References to chapter IV

1. Traxler, M.F., et al., *Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome.* mBio, 2013. **4**(4).
2. Lee, J., M. Simurdiak, and H. Zhao, *Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation.* J Biol Chem, 2005. **280**(44): p. 36719-27.
3. Lee, J. and H. Zhao, *Mechanistic studies on the conversion of arylamines into arylnitro compounds by aminopyrrolnitrin oxygenase: identification of intermediates and kinetic studies.* Angew Chem Int Ed Engl, 2006. **45**(4): p. 622-5.

Summary in French

Avant-propos

Ma thèse s'inscrit dans le cadre du projet IMPACT Biomolécules de l'I-site (« Initiatives-Science - Innovation - Territoires - Economie ») Lorraine Université d'Excellence (LUE). Ce projet a pour but d'identifier et de générer de nouvelles biomolécules présentant des propriétés anti-inflammatoires, anti-prolifératives et antioxydantes pour diverses applications en particulier dans le domaine médical. Il couvre des axes allant de la découverte de produits naturels, au criblage de bioactivités contre des cibles spécifiques et à la fonctionnalisation pour améliorer le potentiel des molécules présentant les activités les plus prometteuses. Cela inclut également des études de formulation/délivrance (encapsulation et vectorisation) pour faciliter la transition vers l'application industrielle.

Mon projet de thèse avait pour but de générer de nouvelles biomolécules polycétidiques par des approches de biologie de synthèse (ingénierie génétique et mutasynthèse) en utilisant comme modèle la voie de biosynthèse des stambomycines, une famille de macrolides produite par la bactérie *Streptomyces ambofaciens* ATCC23877 et présentant des activités antibactériennes et anti-tumorales prometteuses.

Introduction

La recherche de nouvelles molécules bioactives est plus que jamais un sujet d'actualité et notamment dans le domaine de la lutte contre les microorganismes pathogènes. En effet, la résistance aux antimicrobiens représente aujourd'hui une menace majeure en santé publique pouvant conduire à une incapacité à lutter contre les maladies infectieuses. Une projection place l'antibiorésistance au 1er rang des causes de mortalité à l'horizon 2050 (avec 10 millions de décès estimés) si les mesures de lutte s'avèrent inefficaces [1]. Outre l'impact sur la santé, le coût économique lié à l'antibiorésistance a été estimé par le Centre européen de prévention et de contrôle des maladies (ECDC) à près de 1,5 milliard par an au cours de la dernière décennie. Le besoin de nouvelles molécules est également prégnant dans d'autres domaines liés à la santé tels les cancers. En effet, ces derniers représentent l'une des causes majeures de décès dans les pays développés (seconde cause la plus importante de décès aux Etats-Unis). La Société Américaine des Cancers (ACS) estime

que 50% des hommes et 30% des femmes développeront un cancer au cours de leur vie. La découverte de nouveaux agents anti-infectieux and antiprolifératifs représente donc aujourd'hui un défi majeur.

Les polycétides bactériens constituent une source très importante de molécules d'origine microbienne présentant des applications importantes en thérapie, notamment anti-infectieuse et anticancéreuse [2] (**Figure 1**).

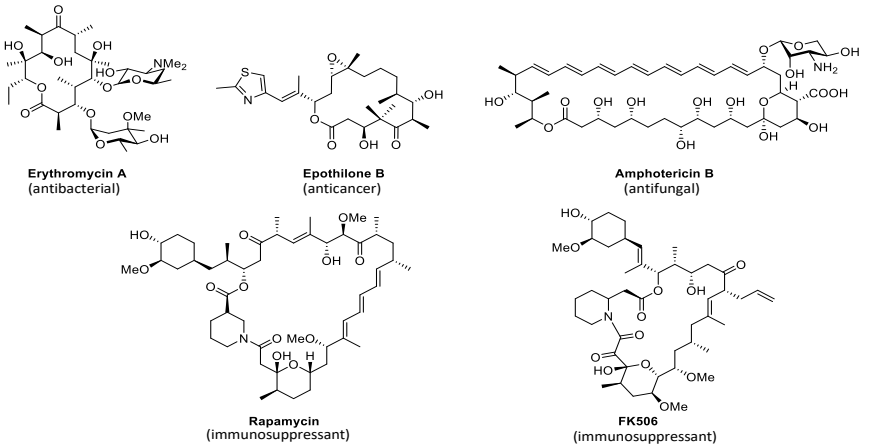


Figure 1 : Exemples de molécules polycétidiques d'origine bactérienne utilisées en thérapie

Cependant, leurs structures doivent souvent être optimisées avant de pouvoir être utilisées comme molécules thérapeutiques [3]. Une approche prometteuse pour obtenir de tels analogues consiste à modifier génétiquement les voies de biosynthèse de ces métabolites et en particulier les polycétides synthases (PKS), les complexes multienzymatiques responsables de l'assemblage du squelette de ces molécules. Les enzymes « post-PKS » qui modifient chimiquement les structures polycétidiques produites par les PKS et dont l'action est souvent nécessaire pour la bioactivité des molécules produites, représentent également des cibles intéressantes.

Les stambomycines (A-F) correspondent à une famille de macrolides « géants » (elles possèdent un noyau macrolactone de 51 atomes) produits par la bactérie *Streptomyces ambofaciens* ATCC23877 (**Figure 2**) et identifiées par une approche de *genome mining* [4]. Ces molécules possèdent des propriétés antibactériennes et surtout antitumorales prometteuses contre de multiples lignées cellulaires cancéreuses.

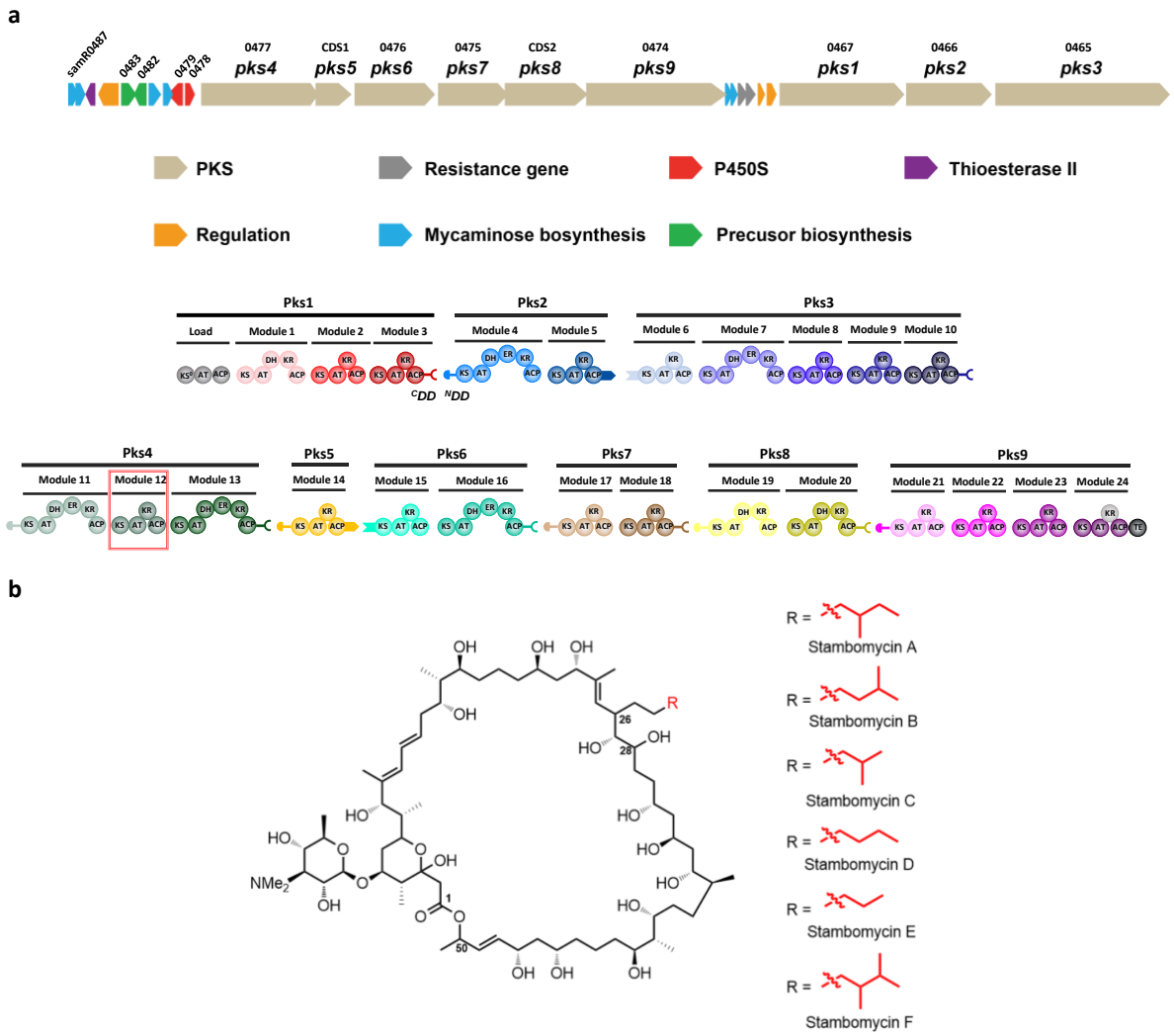


Figure 2 : a. Organisation génétique du cluster de gènes de PKS de *S. ambofaciens* ATCC23877 responsable de la synthèse des stambomycines et organisation des modules et domaines catalytiques. Les domaines de docking (DD) permettant l'interaction spécifique entre les sous-unités sont également représentés. Le domaine acyl-transférase (AT) du module 12 de la PKS (encadré rouge) recrute six unités d'extension alternatives, donnant naissance aux différentes formes de stambomycines ; **b.** Structure des stambomycines A F.

Mon travail de thèse avait comme objectif de générer des dérivés des stambomycines dans l'optique d'obtenir à terme des molécules présentant des activités antimicrobiennes ou antitumorales améliorées voire de nouvelles activités ou de nouvelles cibles thérapeutiques. Nous avons donc développé une approche multidisciplinaire de biologie de synthèse. Une première partie de mon travail a visé à générer par ingénierie génétique des stambomycines « miniatures » (ou « pharmacophores minimaux » [5]), c'est-à-dire des dérivés présentant

un noyau macrolactone de taille réduite et d'évaluer leurs activités anticancéreuses et antibactériennes. Une seconde partie de mon travail a consisté à générer des analogues de stambomycine par mutasynthèse. Il s'agissait ici d'obtenir de nouveaux dérivés présentant des chaînes latérales variées à la position C-26 (**Figure 2**), une position ayant une influence significative sur les propriétés antiprolifératives des stambomycines [4].

Résultats

1. Ingénierie des PKS modulaires responsables de la synthèse des stambomycines et synthèse de dérivés présentant une contraction de leur cycle macrolactone

Cette partie des travaux de ma thèse fait l'objet d'un article actuellement en revue dans Nature Communications "Successes, surprises and pitfalls in modular polyketide synthase engineering: generation of ring-contracted stambomycins"(DOI:10.21203/rs.3.rs-222743/v1), pour lequel je suis 1ère auteure.

Les PKS modulaires de type I constituent une véritable chaîne d'assemblage avec une relation directe entre l'organisation des modules et des domaines catalytiques et la structure de la molécule produite. Cette caractéristique est à la base de la prédiction de l'architecture des polycétides à partir de leurs gènes de biosynthèse et des nombreux travaux visant à concevoir et modifier des PKS afin de produire de nouveaux polycétides de manière prédictible. Cependant, malgré des décennies de recherche, les taux de succès restent faibles, les produits visés n'étant pas synthétisés ou avec des rendements très faibles. La stratégie qui semble la plus prometteuse consiste à utiliser les modules ou les sous-unités comme blocs de base pour l'ingénierie, en raison de la forte interdépendance fonctionnelle des domaines catalytiques constituants ces blocs [6, 7].

Les domaines de docking (DD) situés aux extrémités des sous-unités de la PKS permettent une interaction spécifique entre les sous-unités conduisant au produit polycétidique souhaité [8]. En utilisant la PKS modulaire responsable de la synthèse des stambomycines comme modèle, nous avons développé une stratégie visant à modifier les interactions natives entre les sous-unités par une approche d'ingénierie de domaines de docking avec

comme objectif l'obtention de dérivés de stambomycines présentant une réduction du cycle macrolactone (« pharmacophores minimaux » [5]).

La ligne d'assemblage responsable de la synthèse de stambomycines se compose de 9 sous-unités PKS et donc de 8 paires de DD (**Figure 2a**), ce qui en fait un système idéal pour étudier et exploiter in vivo la base moléculaire de la communication inter sous-unité. L'objectif a été de créer une nouvelle interface entre les sous-unités Pks4 et Pks9 (**Figure 2a**) en utilisant une approche combinant la modification des DD situés dans la région C-terminale de Pks4 ou N-terminale de Pks9 et d'un ensemble d'interactions interdomaines clés dans le transfert et l'extension de la chaîne polycétidique.

Le choix de cette jonction alternative reposait sur trois considérations : i) l'analyse des DD a montré que ceux opérant aux jonctions des sous-unités 4/5 et 8/9 appartiennent à la même classe structurelle [8], et qu'il devrait donc y avoir une compatibilité intrinsèque ; ii) le substrat du premier domaine KS de la sous-unité 9 (KS_9) et celui de KS_5 sont similaires sur les premiers carbones, laissant supposer que KS_9 tolérerait le substrat provenant de la sous-unité 4 ; et iii) Pks4 porte le domaine AT_{12} responsable de la variation structurale entre les différentes formes de stambomycines, et donc conserver Pks4 devait permettre de produire des dérivés de stambomycines A-F présentant un noyau lactone de taille réduite.

La nouvelle jonction entre Pks4 et Pks9 a été produite par ingénierie génétique en échangeant l'hélice α3 du DD C-terminal de Pks4 ($^{C}DD_4$) contre celle de $^{C}DD_8$ (mutant « $^{C}DD_4$ helix swap ») ou en substituant des résidus spécifiques de $^{C}DD_4$ considérés comme des déterminants de la spécificité d'interaction [8] par ceux de $^{C}DD_8$ (mutant « $^{C}DD_4$ SDM »), ces deux modifications devant rendre possible en principe l'interaction de $^{C}DD_4$ avec le DD N-terminal de Pks9 ($^{N}DD_9$). Ces modifications ont été effectuées soit en délétant les sous-unités intermédiaires (Pks5–8) pour supprimer les interactions entre Pks4 et Pks5 et entre Pks8 et Pks9, soit en conservant le locus natif Pks5–8.

L'analyse HPLC-MS des extraits de fermentation des mutants ainsi générés montrait que les stambomycines étaient encore produites chez les mutants dans lesquels les sous-unités intermédiaires (Pks5-8) étaient conservées. En revanche, les mutants dans lesquels le locus

Pks5-8 était délété avaient perdu la capacité de produire des stambomycines. Cependant, aucun dérivé de stambomycines à noyau de taille réduite n'a été détecté, mais nous avons mis en évidence une série de produits correspondant à des intermédiaires polycétidiques libérés au niveau des modules 12 et 13 (**Figure 3**).

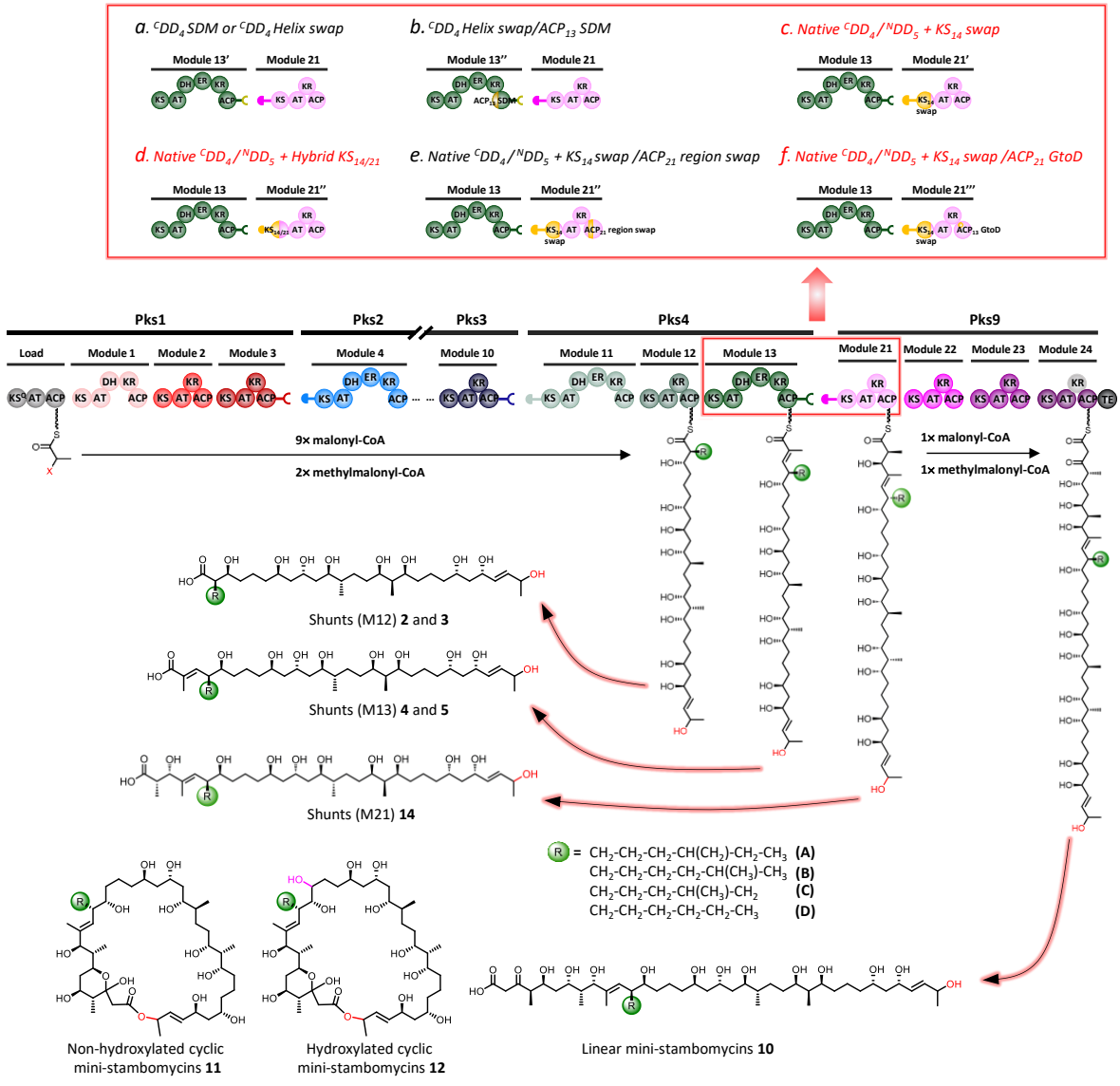


Figure 3 : Stratégies d'ingénierie génétique développées pour obtenir de dérivés de stambomycines présentant un cycle macrolactone de taille réduite. En médaillon figurent les six approches d'ingénierie utilisées, celles en rouge ayant abouti à la synthèse de mini-stambomycines (métabolites 10-12). Le groupe -OH indiqué en rose est introduit par la P450-hydroxylase SamR0478, et celui en rouge, par SamR0479, deux enzymes codées par le cluster stambomycine (**Figure 2a**).

Cette observation indiquait que les jonctions créées étaient probablement non fonctionnelles. Une étude par calorimétrie par titrage isotherme (ITC) des interactions entre des DD recombinants a révélé que la modification 'SDM' était suffisante pour perturber l'interaction native mais insuffisante pour induire une nouvelle interaction («$^{C}DD_4$ SDM » n'interagissait pas avec $^{N}DD_5$ ou $^{N}DD_9$). La stratégie « $^{C}DD_4$ helix swap » générait quant à elle une nouvelle interface productive ($^{C}DD_4$ helix swap interagissait avec $^{N}DD_9$ aussi efficacement que $^{C}DD_8$). Néanmoins, la nouvelle interface 'ACP_{13}-$^{C}DD_4$ helix swap/$^{N}DD_9$-KS_{21}' n'a pas fonctionné in vivo. Ces observations ont renforcé l'idée que la paire de domaines ACP_n/KS_{n+1} devait être prise en compte lors de l'ingénierie des interfaces car cette interaction protéine-protéine contribue à l'efficacité de transfert de la chaîne polycétidique en formation.

Nous avons donc tenté d'améliorer la compatibilité entre ACP_{13}/KS_{21} ou de maintenir la jonction native ACP_{13}/KS_{14} pour faciliter la translocation de la chaîne polycétidique. Pour cela i) l'hélice α1 de ACP_{13} a été modifiée pour correspondre à celle de ACP_{20} (le partenaire natif de KS_{21}) [9]; et, ii) $^{N}DD_9$ et le premier domaine (KS_{21}) ont été remplacés par $^{N}DD_5$+KS_{14} en utilisant deux sites de jonction différents dans le domaine KS (l'un à l'extrémité du domaine KS [10], l'autre situé au milieu du domaine KS [11]). Cette stratégie a permis de créer trois interfaces modifiées entre les sous-unités 4 et 9 : 'ACP_{13} SDM-$^{C}DD_4$ helix swap/$^{N}DD_9$-KS_{21}', 'ACP_{13}-$^{C}DD_4$/$^{N}DD_5$-KS_{14}' et 'ACP_{13}-$^{C}DD_4$/$^{N}DD_5$-$KS_{14/21}$' (**Figure 3**). Nous avons pu ainsi obtenir des dérivés de stambomycines avec un cycle macrolactone à 37 atomes (mini-stambomycines, **Figure 3**), les stambomycines natives ayant un noyau de 51 atomes.

Toutefois, les mini-stambomycines étaient produites avec un rendement très faible par rapport aux stambomycines natives. De plus, des produits dérivés correspondant aux intermédiaires relargués au niveau des modules 12, 13 ou 21 étaient produits (**Figure 3**, métabolites 2-5 et 14). Nous avons pu démontrer que le domaine thioestérase de type I (TEI) de Pks9 joue un rôle actif dans la libération de ces polycétides. La raison du blocage des intermédiaires en amont de la nouvelle interface Pks4/9 pourrait être due à une très faible efficacité d'extension des chaînes polycétidiques au niveau du premier module de Pks9, liée à une incompatibilité potentielle entre le domaine KS_{14} swap ou le domaine

hybride $KS_{14/21}$ et ACP_{21}. Des modifications au niveau du domaine ACP_{21} afin d'améliorer sa compatibilité avec le domaine KS situé juste en amont n'ont cependant permis d'améliorer les rendements.

Ce travail nous a permis de réaliser la modification la plus spectaculaire rapportée à ce jour d'une chaîne d'assemblage PKS intacte en supprimant un total de sept modules internes. Néanmoins, les faibles rendements des produits issus de cette nouvelle ligne d'assemblage démontrent que même les stratégies d'ingénierie qui ont été « optimisées » sur la base de la littérature ne permettent pas de surmonter les nombreux obstacles intrinsèques pour un fonctionnement efficace des lignes d'assemblage modifiées. En particulier, notre travail a mis en évidence que la faible tolérance au substrat des modules en aval des jonctions introduites peut représenter un obstacle important à la réussite de la biologie de synthèse des PKS.

2. Exploitation de la spécificité large du domaine AT_{12} pour la synthèse d'analogues de stambomycines

Les domaines acyltransférases (AT) des PKS modulaires de type I sont responsables de la sélection des unités d'extension spécifiques à incorporer pour l'assemblage des chaînes polycétidiques. Un domaine AT est généralement très spécifique et reconnait une unité d'extension particulière, typiquement du malonyl ou du méthylmalonyl-CoA [12]. Même si plusieurs unités d'extension inhabituelles sont connues pour être utilisées par certaines PKS modulaires, la voie de biosynthèse des stambomycines est originale dans la mesure où le domaine AT du module 12 permet l'incorporation à la position C-26 de six unités d'extension atypiques (pentyle-/hexyle-/heptyle-/octylmalonyle-CoA), donnant lieu aux différentes formes de stambomycines (stambomycines A–F, **Figure 2b**).

Cette spécificité de substrats intrinsèquement large du domaine AT_{12} ouvrait donc la possibilité de générer de nouveaux analogues de stambomycines par des approches de biosynthèse dirigée par les précurseurs (« Precusor-Directed Biosynthesis », PDB) et de mutasynthèse [13]. Et c'est ce qui a fait l'objet de la seconde partie de mon travail de thèse.

Des travaux récents ont montré que deux enzymes codées par le cluster responsable de la synthèse de stambomycines, SamR0482 et SamR0483 (**Figure 2a**), sont responsables de la formation des six unités d'extension atypiques incorporées par le domaine AT_{12} (**Figure 4**). Ces enzymes SamR0482/SamR0483 sont de plus tolérantes sur la sélection du substrat puisqu'il a été démontré par une approche PDB qu'elles étaient capables de produire des unités d'extension contenant des fonctions azides et alcynes ce qui a conduit à la production de deux nouveaux analogues de stambomycines [14].

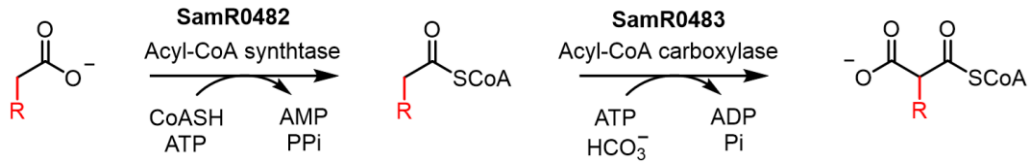


Figure 4 : Voie de synthèse des six unités d'extension atypiques catalysées par SamR0482 et SamR0483 Les deux enzymes permettent la synthèse de six unités d'extension malonyl-CoA atypiques par carboxylation directe d'acides carboxyliques à chaîne moyenne issus du pool cellulaire.

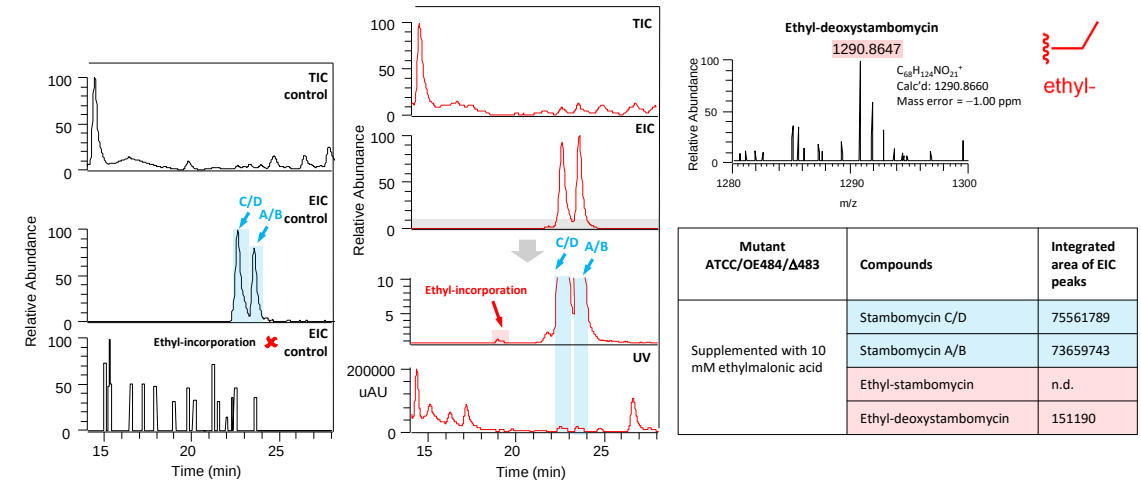
Dans le but de générer de nouveaux analogues de stambomycines par mutasynthèse, la première étape de mon travail a consisté à inactiver le gène samR0483, un mutant délété pour le gène samR0482 étant déjà disponible. Les mutants ont ensuite été cultivés en milieu MP5 (30 °C, sous agitation, 4 jours) puis une analyse par HPLC-MS a été réalisée à partir d'extrait de mycélium (extraction au méthanol).

Les résultats obtenus (chromatogrammes de masse et UV) montrent que la production de stambomycines dans le mutant délété pour le gène samR0483 a été presque complètement abolie (ca. 3% du rendement du type sauvage) ce qui est cohérent avec la perte d'approvisionnement en précurseurs du domaine AT_{12}. La complémentation de ce mutant avec l'allèle sauvage samR0483 se traduisait par un retour de la production des macrolides confirmant le rôle essentiel de SamR0483 dans la synthèse des stambomycines. Le mutant pour le gène samR0482 se comportait essentiellement comme la souche sauvage, c'est-à-dire qu'il était capable de produire les stambomycines bien que le rendement fût réduit à 59 % par rapport à celui de la souche sauvage. L'analyse du génome de *S. ambofaciens* a révélé la présence de plusieurs homologues à samR0482 pouvant expliquer ce résultat.

Le mutant délété pour samR0483 a donc ensuite été cultivé en MP5 dans lequel a été rajouté après 24h de culture différents substrats di-acides : acides éthyle-, allyle-, isopropyle-,

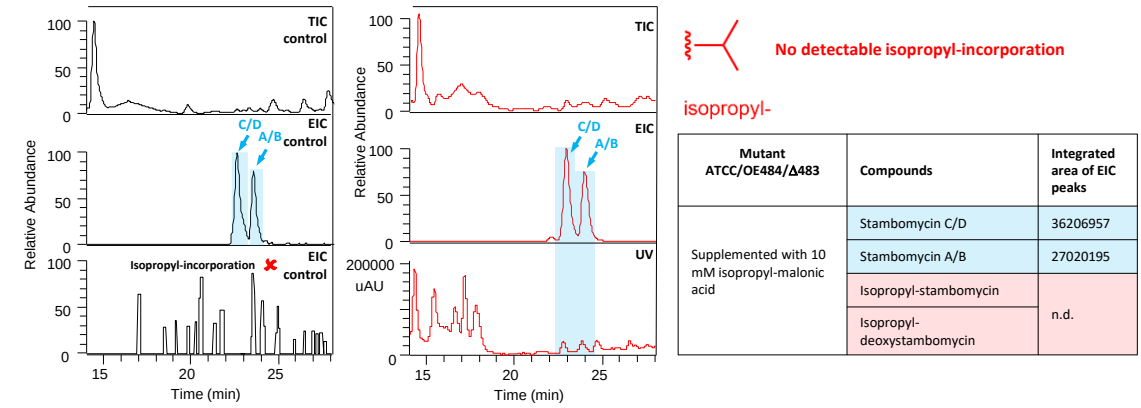
butyle-, benzyle-, octyle-, phénylpropyle-, phénoxypropyle-, thiophène-, thiénylméthyle, propargyle- et 6-bromohexyle-malonique, à une concentration finale de 10 mM. Après 3 jours additionnels de culture, les extraits mycéliens ont été analysés par HPLC-MS. Trois exemples de résultats sont présentés **Figure 5**.

a



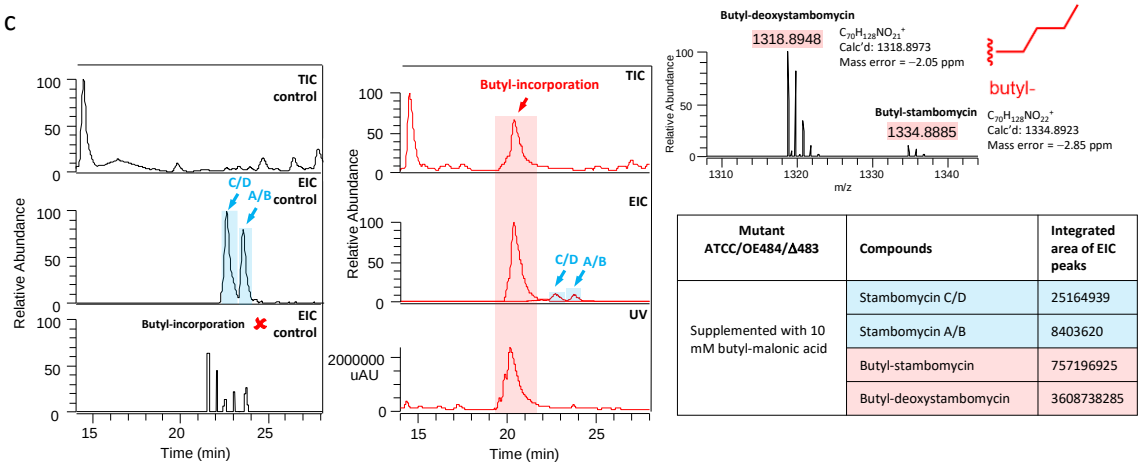
Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM ethylmalonic acid	Stambomycin C/D	75561789
	Stambomycin A/B	73659743
	Ethyl-stambomycin	n.d.
	Ethyl-deoxystambomycin	151190

b



Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM isopropyl-malonic acid	Stambomycin C/D	36206957
	Stambomycin A/B	27020195
	Isopropyl-stambomycin	n.d.
	Isopropyl-deoxystambomycin	

c



Mutant ATCC/OE484/Δ483	Compounds	Integrated area of EIC peaks
Supplemented with 10 mM butyl-malonic acid	Stambomycin C/D	25164939
	Stambomycin A/B	8403620
	Butyl-stambomycin	757196925
	Butyl-deoxystambomycin	3608738285

Figure 5 : Analyse par spectrométrie de masse et détection UV d'extraits bruts de la souche mutante délétée pour le gène samR0483 cultivée en présence d'acide éthyle-, isopropyle-, ou butyle-malonique. Dans chaque cas sont indiqués : le TIC, l'EIC et l'UV238 des extraits, le spectre de masse des analogues de stambomycine correspondants, et les niveaux de production des analogues et des stambomycines natives A-D basés sur l'aire intégrée des pics EIC (moyenne de deux fermentations indépendantes). La mention "n.d." indique les composés qui n'ont pas été détectés.

Bien que l'incorporation d'acides isopropyle-, phényle-, thiophène-, thiénylméthyle-, and propargyle-malonique n'ait pas été observée, des analogues de stambomycines, c'est-à-dire des éthyle-, allyle-, butyle, benzyle-, octyle-, phénoxypropyle-, 6 bromohexyle-stambomycines, ont été obtenus (même si nous ne pouvons pas exclure que l'octyle-stambomycine corresponde en fait à la stambomycine F).

En fait, sur les cinq incorporations réussies (éthyle-, allyle-, butyle-, benzyle- et phénoxypropyle-), les analogues de stambomycines sous une forme dépourvue d'hydroxylation en C-28, appelée « désoxystambomycine », ont été les principaux, voir les seuls, produits obtenus (aucune molécule d'éthyle-stambomycine n'a été détectée), représentant 80-90% des rendements totaux. Au contraire, la 6-bromohexyle-stambomycine est le seul analogue identifié lors des expériences de supplémentation avec du 6-bromohexyle.

Le succès de cette approche de mutasynthèse démontre que la protéine MatB de *S. ambofaciens* (MatB_Sa), qui présente 90% d'identité au niveau aa avec la malonyl-CoA synthétase MatB de Streptomyces coelicolor, une enzyme capable d'activer une large gamme de substrats sous forme d'acyl-CoA [15], possède également une spécificité large permettant d'activer des dérivés exogènes du malonate.

Toutefois, la production de stambomycines était également observée dans toutes les expériences de mutasynthèse indiquant que la compétition entre les unités d'extension natives et non natives existait toujours. Par conséquent, l'avantage théorique de l'approche par mutasynthèse n'était pas évident par rapport à l'approche PDB. Nous avons donc comparé les deux approches. Pour les expériences de PDB, nous avons utilisé la souche sauvage de *S. ambofaciens* ATCC23877 cultivée en présence de monoacides (acides allyle-, butyle-, and benzyle-acétiques) et pour la mutasynthèse la souche mutante samR0483 cultivée en présence de monoacides (acides allyle-, butyle-, and benzyle-maloniques).

Les résultats obtenus ont montré que l'acide butyle-acétique a été activé par la voie SamR0482/SamR0483, la souche sauvage produisant des butyle-stambomycines. En revanche, aucune incorporation d'acide allyle-acétique et d'acide benzyle-acétique n'a été observé. De plus, le rendement des analogues butyle entre la souche sauvage et le mutant samR0483 montrait que l'approche de mutasynthèse est plus efficace que celle basée sur la PDB.

3. Identification de nouveaux dérivés de desferrioxamines

Lors de l'analyse par HPLC-MS des extraits obtenus à partir des souches mutantes générées pour produire des mini-stambomycines, deux séries de métabolites apparentés ont été identifiées avec des masses de 687.5013, 701.5171, 715.5303, 729.5466, 743.5624 pour la première série et des masses de 717.4736, 731.4901, 745.5048, 759.5206, 773.5753 pour la deuxième série.

Ces deux séries de molécules ont été appelées MINI et MINI+30 car elles sont séparées l'une de l'autre par 30 unités de masse. Dans chacune de ces séries, les molécules diffèrent séquentiellement l'une de l'autre par "+14". Initialement, nous avons émis l'hypothèse que les molécules MINI correspondaient aux produits polycétidiques libérés au niveau de la sous-unité Pks4, suivi par la perte d'une molécule d'eau (soit par déshydratation soit par cyclisation). Cependant, ces composés ont également été identifiés dans la souche de type sauvage, bien qu'à des niveaux faibles, ainsi que dans une souche mutante de *S. ambofaciens* dépourvue de la première sous-unité PKS de la stambomycine et dans le mutant utilisé pour l'approche de mutasynthèse. Par conséquent, l'ensemble de ces données indiquait que ces molécules appartenaient à un groupe de nouveaux composés sans lien avec les stambomycines.

Deux autres groupes de pics apparentés ont également été observés lors des analyses par HPLC-MS avec des masses de 903.5765, 917.5923, 931.6079, 945.6235, 959.6350, 973.6550, 987.6705 et 1103.6931, 1117.7086, 1131. 7421, 1145.7396, 1159.7555, 1173.7711, 1187.7870, qui diffèrent séquentiellement des MINI+30 par +200 pour les molécules du 1er groupe et +200+200 pour celles du 2ème groupe. Ces composés ont été annotés MINI+200+30 et MINI+200+200+30. Les sept membres de chaque série de masse

diffèrent séquentiellement par +14. De plus, le temps de rétention pour chaque groupe MINI/MINI+30/MINI+200+30/MINI+200+200+30 était proche. Cela nous a donc conduit à proposer qu'ils étaient tous membres de la même famille de métabolites.

Grâce à la plateforme GNPS (Global Natural Products Social Molecular Networking) [16], un réseau moléculaire basé sur les spectres de fragmentation de l'ensemble de ces composés a été généré. L'analyse de ce réseau a révélé que la série de composés MINI (représentée par des points bleus, **Figure 6**) correspondait à des sidérophores acyle-desferrioxamines (acyle-DFO) et plus précisément à des C9 à C15 acyle-DFO. Les formes chélatant du fer correspondantes de ces molécules, appelées ferrioxamines (acyle-FO, points jaunes **Figure 6**), ont été observées. Nous avons également mis en évidence la bisucabérine (un sidérophore à base de N-hydroxy-N-succinyl diamine (HSD)), la DFO-E, la DFO-D2, une forme de DFO-B chélatant un ion aluminium (Al) (points verts, **Figure 6**), ainsi que la Desf-07 (point turquoise).

Les composés MINI+30, MINI+200+30 et MINI+200+200+30 étaient également détectés dans ce réseau (ils sont représentés respectivement en orange, rose et rouge avec des contours noirs, **Figure 6**). Cela nous a permis de confirmer qu'ils étaient structurellement liés et qu'ils appartenaient à la même famille de métabolites. Toutefois, même s'ils sont connectés aux DFO et acyle-DFO, aucune identification n'était proposée à partir de la base de données GNPS, indiquant que nous avions identifiés trois nouvelles séries de dérivés de DFO. Nous les avons renommés respectivement acyle-DFO+30, acyle-DFO+200+30 et acyle-DFO+200+200+30.

L'analyse des profils de MS/MS de ces composés en comparaison avec ceux décrits dans la littérature nous a alors permis de proposer une structure pour chacun d'eux (**Figure 7**) et de définir l'origine des "+200" et "+30" en masse. Le "+200" correspondrait à l'incorporation d'une unité de N-hydroxy-N-succinylecadavérine (HSC) supplémentaire, une unité d'extension classique des DFO, tandis que le "+30" proviendrait de deux modifications possibles du groupe amine terminal (**Figure 7**).

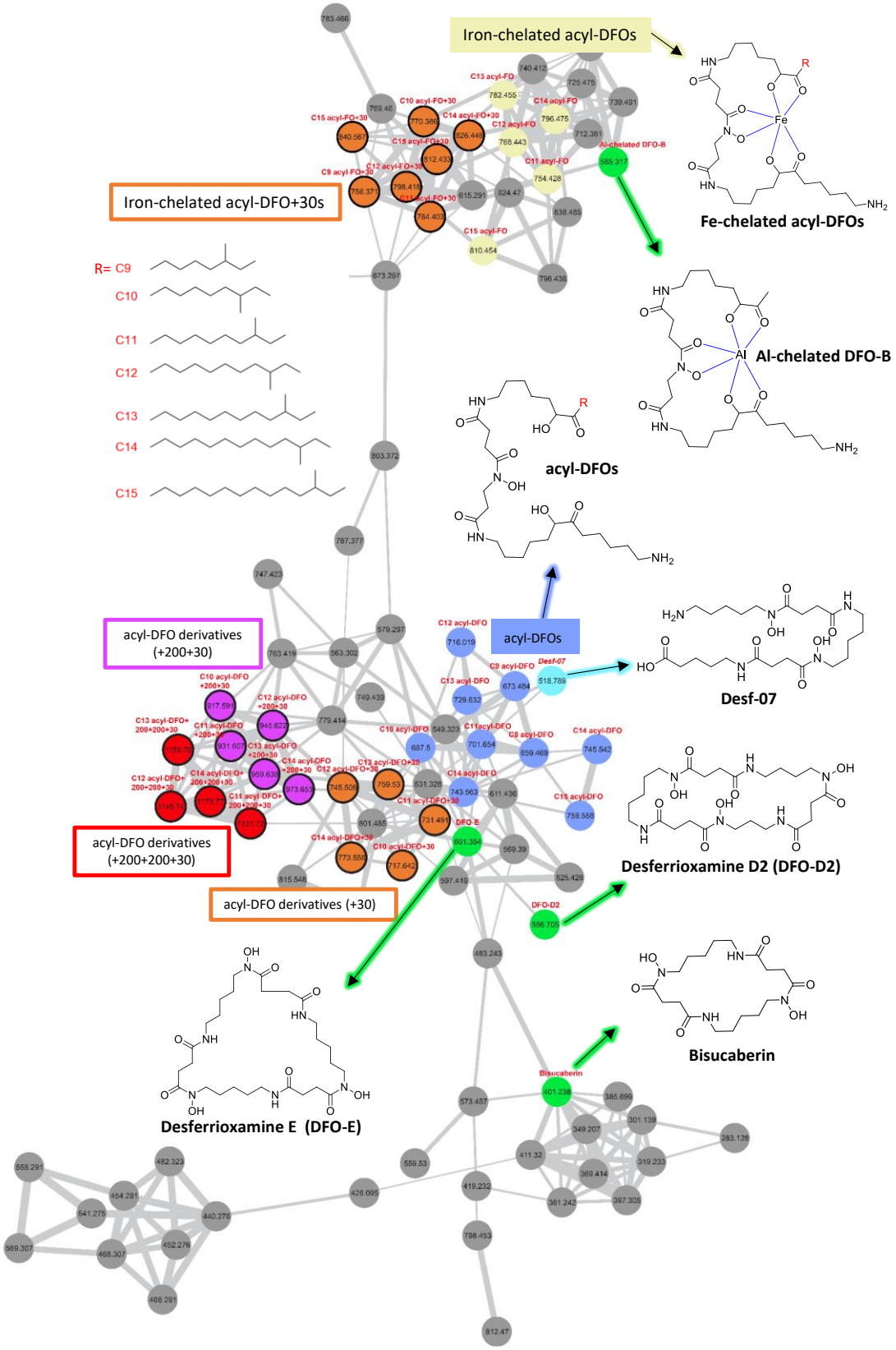

Iron-chelated acyl-DFOs
Iron-chelated acyl-DFO+30s
R= C9
C10
C11
C12
C13
C14
C15
Fe-chelated acyl-DFOs
Al-chelated DFO-B
acyl-DFOs
acyl-DFO derivatives (+200+30)
acyl-DFO derivatives (+200+200+30)
acyl-DFO derivatives (+30)
Desf-07
Desferrioxamine D2 (DFO-D2)
Bisucaberin
Desferrioxamine E (DFO-E)

Figure 6 : Annotation du réseau obtenu lors de l'analyse avec GNPS. La masse protonée en Da de chaque point est indiquée au centre dans ce réseau. Les noms de tous les composés connus sont donnés en haut de chaque point, et leurs structures sont indiquées. Les autres membres des acyle-DFO+30 (C9 et C15), acyle-DFO+200+30 (C9 et C15) et acyle-DFO+200+200+30 (C9, C10 et C15) n'ont pas été présentés dans le réseau, en raison de leurs densités ioniques relativement faibles et du manque de fragments MS/MS disponibles.

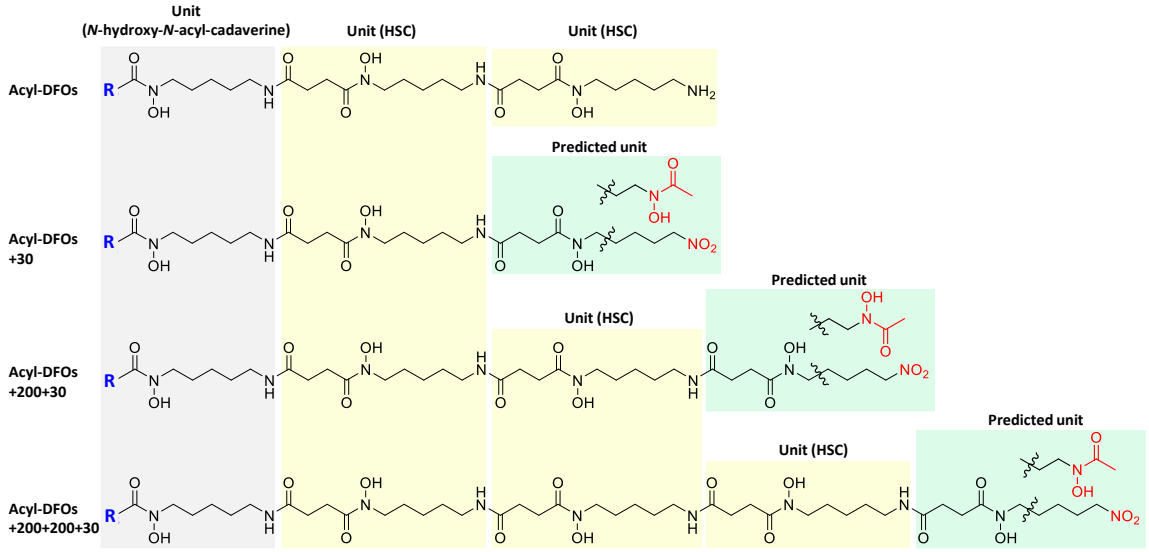


Figure 7 : Comparaison de la composition en monomères des acyle-DFO+30, acyle-DFO+200+30 et acyle-DFO+200+200+30. La structure des acyle-DFO est basée sur la référence [17], tandis que les structures des dérivés supplémentaires ont été prédites sur la base des modèles de fragmentation MS/MS. Les différentes unités représentées dans des couleurs différentes sont reliées entre elles par des liaisons amides pour former les différentes DFO. Le +200 fait référence à l'incorporation d'une unité HSC supplémentaire, tandis que le +30 est prédit comme provenant de l'une des deux modifications alternatives du groupe amine terminal.

Bien que les structures des nouvelles acyle-DFO restent à déterminer, nous proposons deux hypothèses alternatives pour expliquer l'origine du +30 : i) la transformation du groupe amine terminal (-NH2) en un groupe nitro (-NO2), se produisant pendant l'analyse LC-ESI-MS. En effet, pendant l'étape d'ionisation, des réactions électrochimiques peuvent se produire à l'interface liquide-métal sur le capillaire, conduisant à des espèces réactives de l'oxygène (par exemple -OOH et H2O2) capables d'oxyder l'amine [18, 19]. Le NO2 pourrait également provenir de l'oxydation enzymatique du NH2 catalysée par la protéine SAM23877_8801, une oxygénase non hémique [2Fe-2S] présentant une identité de séquence d'acides aminés de 39 % avec PrnD. PrnD est une N-oxygénase de type Rieske impliquée dans la formation d'un groupe nitro à partir d'un groupe amine lors de la

biosynthèse de l'antibiotique antifongique pyrrolnitrine chez Pseudomonas fluorescens [20, 21] ; et, ii) l'utilisation d'un bloc de construction alternatif dérivé du 1,3-diaminopropane pendant l'assemblage du composé.

L'analyse du génome de *S. ambofaciens* ATCC23877 a révélé la présence de deux loci homologues au locus responsable de la synthèse de schizokénine chez la bactérie *Nostoc sp*. PCC7120. En effet, l'un des loci contient des gènes homologues à all0392 (SAM23877_2102, 43% d'identité au niveau de la séquence AA) et all0395 (SAM23877_2103, 39%) tandis que l'autre comprend des homologues de all0396 (SAM23877_5502, 39%), all0393 (SAM23877_5504, 37%) et all0394 (SAM23877_5503, 40%)/ou all0390 (25%). Néanmoins, ces loci ne contiennent qu'un seul homologue (SAM23877_5503) des deux gènes (all0394 et all0390) codant les enzymes NIS (NRPS-independent siderophore) nécessaires à la biosynthèse de la schizokénine. Cette observation explique probablement l'absence de détection de la schizokénine dans les cultures des différentes souches de *S. ambofaciens* analysées. La synthèse de schizokénine implique une unité de du 1,3-diaminopropane. Si +30 correspond à l'incorporation de 1,3-diaminopropane, un cross-talk entre deux voies de biosynthèse, la voie Des responsable de la synthèse de la desferrioxamine et celle responsable de la synthèse de schizokénine, pourrait être à l'origine des trois nouvelles séries de DFO (acyle-DFO+30, acyle-DFO+200+30 et acyle-DFO+200+200+30) identifiées chez *S. ambofaciens* (**Figure 8**).

Dans ce scénario, DesABC serait responsable de la génération des unités à base de cadavérine HSC et N-hydroxy-N-acyl-cadavérine, tandis que le couple d'enzymes SAM23877_5502 /SAM23877_2103 ou SAM23877_5502/DesA serait responsable de la production de 1,3-diaminopropane. Le 1,3-diaminopropane serait alors N-hydroxylé par DesB et SAM23877_2102 (ou N-hydroxylé deux fois par l'une des deux enzymes) sur ses deux extrémités amines, puis N-acétylé et N-succinylé par DesC et SAM23877_5504 (ou par l'une des deux), respectivement. Enfin, tous les monomères seraient condensés entre eux par des liaisons amides par les deux NIS synthétases DesD et SAM23877_5503 agissant ensembles (ou alternativement par l'une des deux), aboutissant ainsi à la synthèse des acyle-DFO+30, acyle-DFO+200+30 et acyle-DFO+200+200+30.

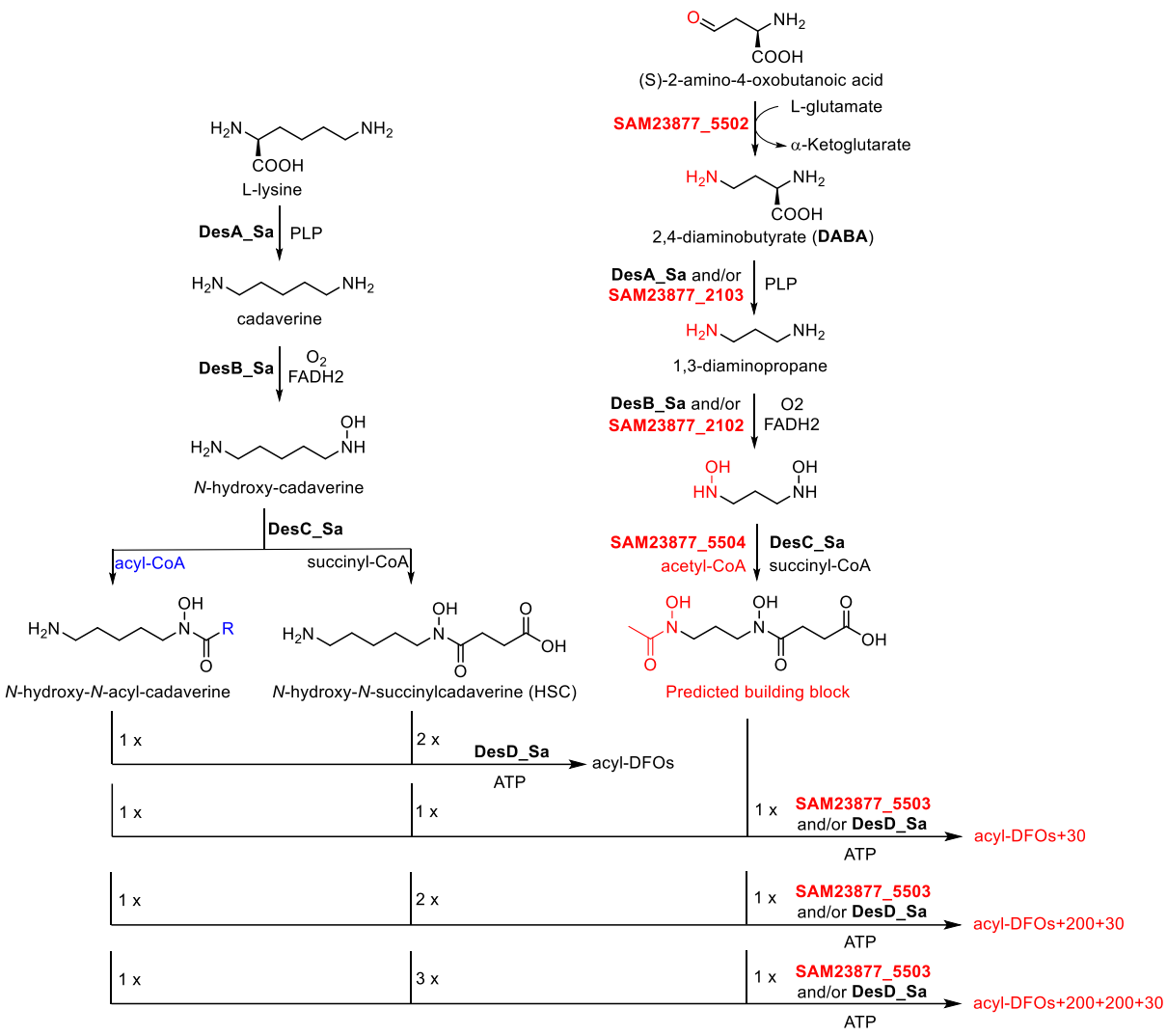


Figure 8 : Proposition de voie de biosynthèse des acyle-DFO+30, acyle-DFO+200+30 et acyle-DFO+200+30 impliquant l'utilisation de 1,3-diaminopropane.

Conclusions

Mon projet de thèse visait à exploiter les connaissances clés sur la biosynthèse modulaire des polycétides et les approches de biologie de synthèse pour générer des analogues de polycétides présentant des activités antiprolifératives (antimicrobiennes ou antitumorales) intéressantes. Le système modèle choisi, la PKS responsable de la synthèse des stambomycines, présentait plusieurs caractéristiques intéressantes à cet égard : i/ une taille inhabituellement grande pour un macrolide, ce qui en fait une cible attrayante pour la réduction du cycle lactone afin d'obtenir des dérivés de taille réduite (des mini-

pharmacophores) et ii/ une tolérance intrinsèque au substrat au niveau domaine acyltransférase du module 12 (AT_{12}) et de tous les modules en aval qui se traduit par la synthèse des formes A–F des stambomycines.

Concernant l'obtention de dérivés de stambomycines à noyau lactone contracté, nous avons généré toute une série de mutants et cela a finalement abouti à l'identification des produits ciblés. En effet, la stratégie basée sur la définition alternative du module (c-à-d des mutants dans lesquels l'interface clé ACP_n/KS_{n+1} au niveau de la nouvelle jonction est préservée) a été couronnée de succès et a permis de générer des mini-stambomycines avec un cycle lactone de 37 atomes. Toutefois, les rendements étaient faibles. La connaissance des systèmes de PKS modulaires nécessite donc encore des approfondissements malgré une littérature déjà abondante. Néanmoins, ce travail a mis en évidence de nouvelles caractéristiques des systèmes PKS hybrides qui pourront servir à l'avenir pour améliorer leur fonctionnement, notamment sur la spécificité du substrat polycétidique des domaines/modules en aval des interfaces modifiées. Ce type d'approche couplé à des analyses bioinformatiques devraient permettre de mieux comprendre les schémas naturels d'évolution des PKS cis-AT (famille à laquelle appartient le complexe de PKS impliqué dans la synthèse des stambomycines), et ainsi de développer des stratégies de fusion plus efficaces et de déployer des méthodes d'évolution dirigée pour surmonter les limites de la manipulation rationnelle de ces systèmes.

La seconde partie de ma thèse a porté sur le développement d'une approche de mutasynthèse, basée sur la flexibilité de reconnaissance d'unités d'extension du domaine AT du module 12 (AT_{12}). Grâce à l'inactivation de l'une des deux enzymes catalysant la biosynthèse des six unités d'extension natives reconnues par AT_{12}, nous avons démontré que la souche mutante obtenue est capable de synthétiser 6 nouvelles stambomycines modifiées en position C-26 via l'incorporation d'unités d'extension non natives dérivées de diacides disponibles dans le commerce. Ce résultat a permis de montrer que l'acyl-CoA synthétase native MatB de *S. ambofaciens* (MatB_Sa) était capable d'activer les diacides utilisés sous forme de thioesters CoA, démontrant une tolérance intrinsèque de cette enzyme. Nous avons également montré que la protéine MatB de Streptomyces cinnamonensis améliorait l'incorporation de certains analogues. Ainsi, ce kit d'outils

composé d'une AT à large spectre et d'enzymes MatB peut être utile pour générer des dérivés modifiés d'autres polycétides. La nature de la chaîne latérale en position C-26 a un impact direct sur l'activité antitumorale et antibactérienne des stambomycines [4]. Par conséquent, les nouveaux analogues obtenus seront évalués pour leur activités antiprolifératives contre des lignées cellulaires cancéreuses. Une déclaration d'invention a déjà été déposée sur ces composés auprès de la SATT Sayens, en vue du dépôt d'un brevet.

Enfin, de façon tout à fait inattendue, nous avons identifié, notamment lors de l'approche d'ingénierie génétique visant à générer des mini-stambomycines, des composés correspondant à des dérivés de desferrioxamines (DFO), des chélateurs de métaux, en particulier de fer, n'ayant aucun lien direct avec la voie des stambomycines. Ces dérivés correspondent à des acyl-DFO et parmi ces molécules, nous avons pu mettre en évidence des acyle-DFO jamais décrites jusqu'à présent. L'hypothèse que nous favorisons est que ces dérivés résulteraient de l'incorporation d'unités supplémentaires du bloc de construction classique (c-à-d d'unités de N-hydroxy-N-succinylecadavérine), mais aussi d'un monomère dérivé de la voie de la schizokénine présente chez *S. ambofaciens*, le 1,3-diaminopropane. Cela constituerait l'un des rares exemples d'interaction entre voies dans la biosynthèse des sidérophores. Une des perspectives de ce travail sera donc de valider ces hypothèses structurelles en purifiant et en caractérisant structurellement les nouveaux métabolites, ainsi qu'en inactivant des gènes spécifiques dans la voie de la schizokénine pour confirmer leur implication. Étant donné que des DFO étendues (DFO constituées d'unités d'extension additionnelles) se sont récemment révélées capables de chélater les métaux compatibles avec les techniques de tomographie par émission de positrons (tel que le Zr^{4+}) [22, 23], il sera également intéressant d'évaluer les propriétés de chélation des métaux de ces nouvelles acyle-DFO identifiées dans ce travail.

Références bibliographiques

1. de Kraker, M.E., A.J. Stewardson, and S. Harbarth, *Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?* PLoS Med, 2016. **13**(11): p. e1002184.
2. Martin, J.F. and A.L. Demain, *Control of antibiotic biosynthesis.* Microbiological Reviews, 1980. **44**(2): p. 230-251.
3. Weissman, K.J., *Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology.* Natural Product Reports, 2016. **33**(2): p. 203-230.

4. Laureti, L., et al., *Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(15): p. 6258-6263.
5. Ji, H.T., et al., *Minimal pharmacophoric elements and fragment hopping, an approach directed at molecular diversity and isozyme selectivity. Design of selective neuronal nitric oxide synthase inhibitors.* Journal of the American Chemical Society, 2008. **130**(12): p. 3900-3914.
6. Dutta, S., et al., *Structure of a modular polyketide synthase.* Nature, 2014. **510**(7506): p. 512-517.
7. Whicher, J.R., et al., *Structural rearrangements of a polyketide synthase module during its catalytic cycle.* Nature, 2014. **510**(7506): p. 560-564.
8. Broadhurst, R.W., et al., *The structure of docking domains in modular polyketide synthases.* Chemistry & Biology, 2003. **10**(8): p. 723-731.
9. Kapur, S., et al., *Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation.* Proceedings of the National Academy of Sciences, 2012. **109**(11): p. 4110.
10. Yuzawa, S., et al., *Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production.* ACS Synthetic Biology, 2017. **6**(1): p. 139-147.
11. Wlodek, A., et al., *Diversity oriented biosynthesis via accelerated evolution of modular gene clusters.* Nat Commun, 2017. **8**(1): p. 1206.
12. Moore, B.S. and C. Hertweck, *Biosynthesis and attachment of novel bacterial polyketide synthase starter units.* Natural Product Reports, 2002. **19**(1): p. 70-99.
13. Weissman, K.J., *Mutasynthesis – uniting chemistry and genetics for drug discovery.* Trends in Biotechnology, 2007. **25**(4): p. 139-142.
14. Ray, L., et al., *A crotonyl-CoA reductase-carboxylase independent pathway for assembly of unusual alkylmalonyl-CoA polyketide synthase extender units.* Nature Communications, 2016. **7**.
15. Hughes, Amanda J. and A. Keatinge-Clay, *Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Func-tional Showcasing of Streptomyces coelicolor MatB.* Chemistry & Biology, 2011. **18**(2): p. 165-176.
16. Wang, M., et al., *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.* Nature Biotechnology, 2016. **34**(8): p. 828-837.
17. Traxler, M.F., et al., *Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome.* mBio, 2013. **4**(4).
18. Van Berkel, G.J. and V. Kertesz, *Using the electrochemistry of the electrospray ion source.* Anal Chem, 2007. **79**(15): p. 5510-20.
19. Wabner, D. and C. Grambow, *Reactive intermediates during oxindation of water lead dioxide and platinum electrodes.* Journal of Electroanalytical Chemistry and Interfacial Electrochemistry, 1985. **195**(1): p. 95-108.
20. Lee, J., M. Simurdiak, and H. Zhao, *Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation.* J Biol Chem, 2005. **280**(44): p. 36719-27.
21. Lee, J. and H. Zhao, *Mechanistic studies on the conversion of arylamines into arylnitro compounds by aminopyrrolnitrin oxygenase: identification of intermediates and kinetic studies.* Angew Chem Int Ed Engl, 2006. **45**(4): p. 622-5.
22. Guérard, F., et al., *Investigation of Zr(IV) and 89Zr(IV) complexation with hydroxamates: progress towards designing a better chelator than desferrioxamine B for immuno-PET imaging.* Chem Commun (Camb), 2013. **49**(10): p. 1002-4.

23. Patra, M., et al., *An octadentate bifunctional chelating agent for the development of stable zirconium-89 based molecular imaging probes.* Chem Commun (Camb), 2014. **50**(78): p. 11523-5.

Abstract

The polyketide secondary metabolites of bacteria are a rich source of bioactive agents, with notable applications in anti-infective and anti-cancer therapy. However, their structures often need to be optimized in order to tailor their therapeutic and biophysical properties. The 51-membered macrolide stambomycins, among the largest of known polyketides, were recently discovered by genome mining in *Streptomyces ambofaciens* ATCC23877, and notably exhibit promising anti-cancer activity. The family encompasses six members which differ from each other in the alkyl functionality at C-26, due to the alternative choice of extender units by an exceptional acyl transferase domain (AT_{12}) of the modular polyketide synthase (PKS) responsible for synthesizing the stambomycin core. Given their enormous size of the stambomycins and the intrinsic promiscuity of AT_{12}, there is substantial interest in accessing ring-contracted and C-26 substituted derivatives of this compounds which might retain the bioactivity of the parental structures, or exhibit improved or even new properties.

In this work, we have leveraged our current understanding of modular PKS systems to internally contract the stambomycin assembly line, leading to the successful generation, albeit at low yield, of target smaller derivatives (37-membered 'mini-stambomycins'). By careful analysis, we could identify multiple factors contributing to the low titers, information which should inform future engineering strategies. Furthermore, using a mutasynthesis strategy, we were able to exploit the broad specificity of the AT_{12} domain to create six novel C-26 substituted stambomycin analogues. Finally, we unexpectedly identified three series of novel desferrioxamine siderophores produced by *S. ambofaciens.* As a number of key metabolites generated in this work have potential interest for therapeutic applications, they will be targeted for purification, structural characterization and biological evaluation.

Key words: stambomycin, modular PKS engineering, minimal pharmacophores, acyl transferase, desferrioxamine

Résumé

Les polycétides d'origine bactérienne sont une source importante de molécules anti-infectieuses et anticancéreuses d'origine naturelle utilisées en thérapie. Cependant, leurs structures doivent souvent être optimisées afin d'améliorer leurs propriétés thérapeutiques. Les stambomycines, une famille de macrolides parmi les plus grands polycétides connus (elles possèdent un noyau macrolactone à 51 atomes), ont été récemment découvertes par une approche de *genome mining* chez la bactérie *Streptomyces ambofaciens* ATCC23877. Ces molécules présentent une activité anticancéreuse prometteuse. Six formes de stambomycines ont été caractérisées. Elles diffèrent les unes des autres par la fonctionnalité alkyle au niveau de la position C-26 du noyau macrolactone. Cette variabilité est due au choix alternatif d'unités d'extension par un domaine acyltransférase exceptionnel, le domaine AT_{12}, de la polycétide synthase (PKS) modulaire responsable de la synthèse des stambomycines. Étant donné la taille importante du cycle lactone et la promiscuité intrinsèque du domaine AT_{12}, il existe un intérêt substantiel à accéder à des dérivés de stambomycines par réduction de la taille du noyau ou par substitution de la chaine latérale en C-26. Ces dérivés pourraient conserver la bioactivité des structures parentales ou présenter des propriétés améliorées voire nouvelles.

Dans ce travail, nous avons mis à profit notre compréhension actuelle des systèmes de PKS modulaires pour raccourcir en interne la chaîne de production de stambomycines, ce qui a permis de générer avec succès, bien qu'à faible rendement, des dérivés avec un noyau macrolactone de taille réduite (des "mini-stambomycines" présentant un noyau à 37 atomes). Grâce à une analyse minutieuse, nous avons pu identifier les multiples facteurs expliquant les faibles rendements, ce qui permettra d'éclairer les futures stratégies d'ingénierie. En outre, grâce à une stratégie de mutasynthèse, nous avons pu exploiter la large spécificité du domaine AT_{12} pour générer six nouveaux analogues de stambomycine présentant différentes substitutions en position C-26. Enfin, nous avons identifié de manière tout à fait inattendue trois classes de sidérophores étroitement apparentées aux desferrioxamines produites par *S. ambofaciens*. Un certain nombre des métabolites clés générés pendant ce travail constitue une source de nouvelles biomolécules avec des applications thérapeutiques potentielles. Les prochaines étapes consisteront à purifier ces composés, à déterminer leur structure puis à évaluer leurs activités biologiques.

Mots clés : stambomycine, ingénierie PKS modulaire, pharmacophores minimaux, acyltransférase, desferrioxamine