DISS. ETH NO. 28113

Towards Engineering *trans*-Acyltransferase Polyketide Synthases

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

HANNAH ASTRID MINAS
MSc Biology, ETH Zürich
Born on 02.03.1996
Citizen of Zürich, Switzerland and Germany

accepted on the recommendation of

Prof. Dr. Jörn Piel
Prof. Dr. Julia Vorholt
Prof. Dr. Peter Kast

2022

Acknowledgements

First and foremost, I would like to thank **Prof. Jörn Piel** for giving me the opportunity to conduct my doctoral thesis in his lab. You are always supportive of trying unconventional experiments alongside main projects, some of which were key in this thesis. Also, your knowledge on any kind of topic is immense. But while others may dwell on it, your door is always open and thanks to your very down-to-Earth personality, you are approachable at any time. Many thanks also go to my committee members **Prof. Julia Vorholt** and **Prof. Peter Kast** for input and insightful discussions during our meetings.

As my supervisors when I first came into contact with the Piel group during my studies, I would like to thank **Alexander Brachmann** and **Hanna Maciejewska**. You not only taught me that not every PCR necessarily works, but also introduced me to the fascinating world of biosynthetic pathways. For initial guidance and the introduction to the world of PKS recombineering, I am thankful to **Silke Reiter**. I very much appreciate that you took the time to teach me some basics in our few months of overlap in the lab. Further, I would like to thank the entire Piel group for the support and input over the years. It was a pleasure to work with you all in such a warm atmosphere. Special thanks go to **Roy Meoded**. For introducing me to the fun of SNACking, for brainstorming about the weirdest of experimental ideas and for pushing me to "just try it". Without you, a large part of this thesis would not have happened. Many thanks also go to **Stefan Leopold-Messer** for the shared joy in KS trees, impromptu cloning troubleshooting sessions, and an impressive knowledge on PKSs. I would also like to thank **Silke Probst** for support with HPLC experiments and for the company during midnight assays or the writing of this thesis. For help with any kind of synthesis or silica column, I am grateful to **Cora Dieterich**. Thank you for your trust in my non-chemist abilities and understanding my fear of any chemical with a danger sign on it. I am also grateful to **Michael Rust** for being a walking cloning encyclopedia and for ensuring we always had shelter and a poster on conferences. For very detailed protocols and the encouragement to not shorten them too much, I would like to thank **Franziska Hemmerling**. Further, I would like to thank **Amy Fraley** for joint intermediate analysis marathons, frustrated or over-motivated discussions about projects and help with protein structure-related struggles. For conversations about the more bizarre kind of things I would like to thank **Lida Vadakumchery**. Big thanks go to **Tom Scott**, **Amy**, **Stefan**, **Roy**, and **Franzi** for input and corrections on my thesis.

I also had great pleasure working with my student **Sarah**, who contributed important ideas and data to the hybrid project. Thanks also go to my "big" student **Mathijs Mabesoone**, who mastered the initial hurdles of completely switching the field of research quickly and became passionate about cloning and compound isolation faster than anyone I know.

Next, I would like to thank everyone who kept me sane outside of lab. Thanks go to **Franzi**, **Roy**, and **Silke** for running or vitaparcour sessions in the forest in any kind of weather. Many thanks to **Shekhar** for office badminton games and to **Nik** and **Silke** for joining for Kondi. Further, I would like to thank our office shrimpies for existing - your resilience is unparalleled. I would also like to take the opportunity to thank the rest of my office - old and new - for shared "Notvorrat" and shrimp parenting and to **Tom** and **Roy** for support during Covid isolation. Outside of the Piel group, I would like to thank **Marco** for squash sessions and **Tobias** for trips to cubing competitions anywhere between Finland, Spain, and North Macedonia. Thanks also go to **Gilda** and **Josias** for fierce chess games at any given time of the day and to **Bernhard** for stimulating battles in cubing or chess and the motivation to learn random facts or languages.

Last but not least, I would like to thank my family. I am grateful to my equally quiet Opa for making me feel less lost in a world full of people. Thanks also go to my big little sister, for shared laughs and sorrows over all the years, for a basic training in arguing and long walks to talk about plans and secrets. Finally, all of this would not have been possible without my parents. Thank you for introducing me to pipet tips as a toddler, for your endless support, and for your trust in whatever I was doing.

Table of Contents

Abstract

Natural products are chemicals produced by organisms to provide an evolutionary advantage to tackle ecological challenges including the competition against other organisms. Depending on the context, natural products can, for example, provide mediators for symbiotic relationships between or chemical weapons against other organisms. These properties were serendipitously found to be of great use for humans in the fight against illnesses such as bacterial infections. Ever since the efficacy of various natural products was discovered, some already thousands of years ago, people were fascinated by them and aimed to exploit their powers, one important area being antibiotics. After a surge in new discoveries in the middle of the last century, the rise in antibiotic resistances has shifted the focus of research toward new strategies of natural product research and the intertwined field of antibiotic or, more general, drug discovery. Chemists are exploiting total synthesis approaches or derivatize successful lead products in the hope to improve their pharmaceutical properties. Still there is the irrefutable need for new active drug scaffolds and easier methods for the synthesis of drugs. A large class of pharmaceutically relevant compounds are polyketides, which are biosynthetically produced by gigantic enzymatic assembly lines termed polyketide synthases (PKSs). This work centers on type I PKSs, which are comprised of a series of modules which themselves contain different catalytic domains. While each module is involved in the stepwise elongation of an assembly-line bound intermediate by a simple building block, the modification can be deduced from the module's domain architecture. The correlation between a polyketide structure and module architecture has inspired scientists to artificially design and engineer PKSs for the programmed production of bespoke polyketides within bacteria. In this work, we have focused on the evolutionary distinct class of PKSs called *trans*-acyltransferase (*trans*-AT) PKSs, for which thus far no example of a successfully engineered modular system was reported.

Chapter II focuses on the characterization of an unusual module architecture occurring in the lobatamide *trans*-AT PKS. Using deductive reasoning based on the architecture of the biosynthetic assembly line and the final product, we postulate that the starting modules in the lobatamide *trans*-AT PKSs are responsible for the incorporation of an unprecedented, methylated oxime moiety. In biochemical assays, we demonstrate that a module-integrated flavin-dependent monooxygenase domain is capable of oxime formation in assays with various aminoacyl-*N*-acetylcysteamines that were tested with it. Further, we heterologously produce the starting modules and show that the methyltransferase domain in the following module methylates the oxime-bearing intermediate utilizing *S*-adenosyl methionine. Together, these findings expand the biochemical toolbox in *trans*-AT PKSs.

In Chapter III, we take first steps toward engineering *trans*-AT PKSs. Given that the final step of polyketide biosynthesis is the release of the ultimate intermediate from the assembly line, we focus on the dedicated domains responsible for that step. In a series of engineered assembly lines, we exchange the offloading module of the bacillaene *trans*-AT PKS with parts from architecturally related pathways. We also test domains with different offloading mechanisms, which accept chemically distinct penultimate intermediates and test various recombination sites to find a suitable region for fusion between different PKS proteins. We identify a set of basic principles, both genetic and chemical, that must be considered when venturing on *trans*-AT PKS recombineering. These include the chemical properties of the final intermediate and the mode of release.

In Chapter IV, we attempt to study the channeling of non-native biosynthetic intermediates across the first chimeric *trans*-AT PKS assembly lines. We rely on a highly specific protein-protein interaction of an exotic halogenation module from the oocydin *trans*-AT PKS. Using this protein complex as a handle, we introduce a range of non-native PKS modules into the native host and analyze the propagation of the substrate from native to non-native PKS parts. In addition to biosynthetic logic, we use statistical coupling analysis to gain insights into potential fusion sites that minimize the disturbance of interacting

residues. This combined approach leads to the identification of a successful fusion site for the recombineering of *trans*-AT PKSs. Apart from minimal extending modules, we include a set of larger reducing modules and a terminal module to generate hybrid PKSs. We also test the *trans*-AT PKS correlation rule which states that the central KS co-evolves with the upstream domains and thereby recognizes specific incoming intermediate. In engineered strains with truncated modules, we confront KSs with non-cognate intermediates and show that intermediate transfer is indeed impeded.

This thesis contributes to the understanding and utilization of the most complex natural product family currently known. By developing and validating a set of engineering rules, the work provides first insights into how *trans*-AT PKSs can be reprogrammed in a rational way for the bacterial production of designed, pharmaceutically relevant polyketides.

Zusammenfassung

Naturstoffe sind chemische Verbindungen, die von Organismen produziert werden, um ökologische Herausforderungen wie zum Beispiel den Wettbewerb um Ressourcen mit anderen Organismen zu bewältigen. Je nach Kontext sind Naturstoffe Informationsträger in symbiotischen Beziehungen, oder dienen als chemische Waffen gegen andere Organismen. Eher zufällig wurde festgestellt, dass diese Eigenschaften auch für den Menschen bei der Bekämpfung von Krankheiten wie bakteriellen Infektionen von grossem Nutzen sein können. Seit der Entdeckung der Heilkraft von Naturstoffen vor Tausenden von Jahren, war die Menschheit von ihnen fasziniert und versuchte, sie in der Form von Antibiotika zu nutzen. Nach einer anfänglichen Welle neu entdeckter Naturstoffe Mitte letzten Jahrhunderts, hat die Entwicklung von Resistenzen den Forschungsschwerpunkt auf neue Strategien in der Naturstoffforschung und das damit verbundene Gebiet der Antibiotika- und Arzneimittelentdeckung verlagert. Chemiker nutzen Totalsyntheseansätze oder derivatisieren erfolgreiche Leitstrukturen in der Hoffnung, deren pharmazeutische Eigenschaften zu verbessern. Nach wie vor besteht ein enormer Bedarf an neuen Wirkstoffgerüsten und einfacheren Methoden für die Synthese von Arzneimitteln. Die Polyketide bilden eine grosse Klasse pharmazeutisch relevanter Verbindungen, welche biosynthetisch von gigantischen enzymatischen Proteinkonstrukten, den Polyketidsynthasen (PKS), hergestellt werden. Diese Arbeit konzentriert sich auf Typ I PKS, die aus einer Reihe von Modulen bestehen, die ihrerseits verschiedene katalytische Domänen enthalten. Während jedes Modul an der schrittweisen Verlängerung eines enzymgebundenen Zwischenprodukts durch einen einfachen Baustein beteiligt ist, kann die eingeführte chemische Modifikation aus der Domänenarchitektur des Moduls abgeleitet werden. Diese Korrelation zwischen Polyketidstruktur und Modularchitektur hat das Interesse an künstlich hergestellten PKS für die programmierbare Produktion von Polyketiden in Bakterien geweckt. In dieser Arbeit haben wir uns auf die evolutionär von anderen PKS unterscheidbare Klasse von *trans*-Acyltransferase (*trans*-AT) PKS konzentriert, von welchen bisher keine umgebauten modularen Systeme beschrieben worden sind.

Kapitel II konzentriert sich auf die Charakterisierung einer neuartigen Modularchitektur in der Lobatamide *trans*-AT PKS. Mithilfe der enzymatischen Modularchitektur des biosynthetischen 'Fliessbands' (*engl.: Assembly Line*) und der Struktur des Endprodukts postulieren wir, dass die Startmodule in der Lobatamide *trans*-AT-PKS für den Einbau einer beispiellosen, methylierten Oximeinheit verantwortlich sind. In biochemischen Versuchen zeigen wir, dass eine modulintegrierte Flavin-abhängige Monooxygenase verschiedene Aminoacyl-*N*-acetylcysteamine zu Oximen umsetzt. Darüber hinaus stellen wir die Startmodule heterolog her und zeigen, dass die Methyltransferasedomäne im nachfolgenden Modul das Oxim-Zwischenprodukt mithilfe von *S*-Adenosylmethionin methyliert. Diese Erkenntnisse erweitern das bekannte biochemische Repertoire für *trans*-AT PKS-Biosynthesewege.

In Kapitel III unternehmen wir erste Schritte zur Herstellung neu konstruierter *trans*-AT PKS. Da in der Polyketidbiosynthese im letzten Schritt jeweils das finale Intermediat von der *Assembly Line* abgespalten wird, konzentrieren wir uns zuerst auf die dafür verantwortlichen Module. Wir konstruieren verschiedene *Assembly Lines*, indem wir die Freisetzungsmodule der Bacillaene *trans*-AT-PKS durch architektonisch ähnliche Gegenstücke aus anderen Biosynthesewegen austauschen. Wir testen Domänen mit unterschiedlichen Freisetzungsmechanismen, die chemisch unterschiedliche Intermediate akzeptieren. Ausserdem testen wir verschiedene Rekombinationsstellen, um eine geeignete Region für die Fusion von nicht-nativen PKS-Proteinen zu finden. So identifizieren wir eine Reihe von Grundprinzipien, die bei der Rekombination von *trans*-AT PKS berücksichtigt werden müssen. Diese schliessen die chemischen Eigenschaften des letzten Biosyntheseintermediats sowie den Freisetzungsmechanismus ein.

In Kapitel IV untersuchen wir die Weitergabe nicht nativer Biosynthese-Zwischenprodukte zwischen chimären *trans*-AT PKS *Assembly Lines*. Hierzu machen wir uns die hochspezifische Protein-Protein-Wechselwirkung des exotischen Halogenierungsmoduls in der Oocydin *trans*-AT-PKS zunutze. Diesen Proteinkomplex nutzen wir als Werkzeug, um eine Reihe nicht-nativer PKS-Module in den natürlichen Produzenten einzuführen und analysieren die Substratweitergabe von nativen zu nicht-nativen PKS-Abschnitten. Zusätzlich zur biosynthetischen Logik nutzen wir die statistische Kopplungsanalyse, um potenzielle Fusionsstellen zu identifizieren, bei welcher die Interaktion von Aminosäureresten minimal gestört wird. Dieser kombinierte Ansatz führt zur Identifizierung einer geeigneten Fusionsstelle für die geplante Rekombination von *trans*-AT PKS. Neben einer Reihe an minimalen Verlängerungsmodulen testen wir auch grössere Module und ein Freisetzungsmodul, um Hybrid-PKS zu generieren. Darüber hinaus prüfen wir die *trans*-AT PKS-Korrelationsregel, die besagt, dass die zentrale Ketosynthase mit den davorliegenden Domänen ko-evolviert und dadurch eine Spezifität für das akzeptierte Zwischenprodukt entwickelt. In genmanipulierten Bakterienstämmen mit verkürzten Modulen konfrontieren wir Ketosynthasedomänen mit unbekannten Zwischenprodukten und zeigen, dass deren Transfer tatsächlich beeinträchtigt ist.

Diese Arbeit leistet einen Beitrag zum Verständnis und der Nutzung der derzeit komplexesten bekannten Naturstoffklasse. Durch die Entwicklung und Validierung erster Regeln für die Herstellung hybrider PKS, liefert die vorliegende Arbeit erste Erkenntnisse darüber, wie *trans*-AT PKS rational für die bakterielle Herstellung von designten, pharmazeutisch relevanten Polyketiden umprogrammiert werden können.

Chapter I

Introduction

1. Introduction

1.1 Natural Products - overview

In any ecological setting, organisms produce chemical compounds to survive, but also to interact. While so-called primary metabolites like lipids, nucleotides, or amino acids are essential for the assembly of membranes, the preservation of genetic information or the assembly of proteins, other compounds have less apparent functions, which only become obvious in the need for survival or symbiosis between different organisms. Originally termed secondary metabolites for their seemingly secondary importance, they were shown to be involved in various processes. Even though many of them are not constantly produced due to high energetic costs, the biosynthesis and thus provision of these metabolites can be set into motion quickly in response to specific environmental stimuli.[1] Examples include the biosynthesis of specific siderophores for improved metal ion uptake during starvation periods[2] or the production and secretion of coloring pigments or fragrances to attract or repel different organisms for pollination or reproduction.[3] Defense against an enemy or competitor by secreting antibiotic compounds provides another example of secondary metabolite-mediated interactions between organisms.[4] With the discovery of more diverse functions of secondary metabolites, they are now commonly referred to as specialized metabolites or natural products to account for their non-secondary functions.[5] The fact that these natural products are involved in interactions between all types of organisms makes them excellent candidates for drug discovery. Some examples are provided in Figure 1A and include the plant-derived antimalarial drug artemisinin (from the plant *Artemisia annua*) and the fungus-derived antibacterial drugs penicillin (from *Penicillium notatum*) and cephalosporin (from *Cephalosporium acremonium*). While first findings in the field of natural product discovery date back to Mesopotamia and Egypt and first drugs were mainly of plant origin or the plants themselves,[6] the so-called "Golden Age of Antibiotics" began many centuries later with Alexander Fleming's discovery of penicillin in 1928.[6b] Even though the subsequent mass production was spurred by historically sad circumstances, it marked the beginning of large-scale production of natural product-derived drugs. The specific targeting of enzymes involved in bacterial cell wall biosynthesis made penicillins an ideal first lead scaffold for novel antibiotics by semi-synthesis. Today, the majority of modern drugs are derived from either natural products directly or their chemically optimized derivatives.[5a]
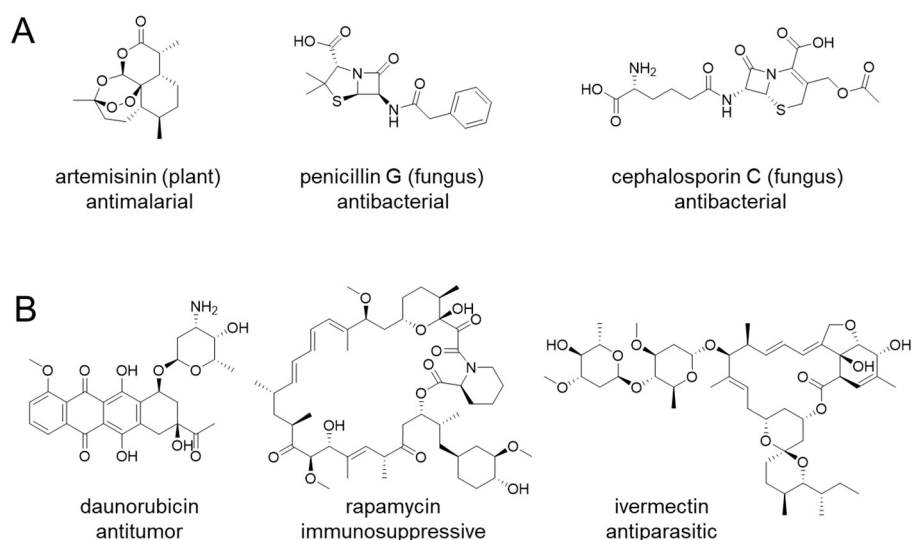

Figure 1. Selected natural products. A) Examples of plant- and fungus-derived natural products. The source is indicated in parentheses and the clinical use below the compound name. B) Examples of bacterially produced natural products.

Over the decades following Fleming's discovery, numerous natural products were discovered in fungal and bacterial sources.[7] Bacteria in particular have proven to be an especially accessible and rich source of bioactive molecules. As easily isolated soil-dwelling producers of antibacterial aminoglycosides and polyketides, *Streptomyces* spp. were among the most investigated species. Some examples of bacterial polyketides include the immunosuppressant rapamycin, antiparasitic drugs such as the semisynthetic ivermectin, and antitumor drugs including daunorubicin (Figure 1B). For an overview of natural product drug leads, see the reviews by Newman and Cragg.[7-8] The Golden Age of antibiotic discovery was short-lived however, with increasing rates of antibiotic rediscovery and the exhaustion of the biosynthetic potential of soil microbes. The issue of low rates of novel scaffold discovery in the decades following the Golden Age was compounded by the emergence of antibiotic-resistant pathogens, arising from the misuse of the antibiotics discovered in previous decades. Multi-drug resistant bacteria such as methicillin-resistant *Staphylococcus aureus* have become prevalent in hospital settings and can only be treated by a limited number of last-resort antibiotics to which natural resistance has not yet arisen. Even though countries have called to reduce unnecessary antibiotic treatments - with some success as demonstrated by a recent urinary sampling study in Chinese children[9] - and to prevent unnecessary prescriptions, we are now in a situation, wherein the need for new antibiotics to replace those that are no longer effective is greater than ever.

1.1.1 Natural product discovery

Classical natural product isolation strategies mainly rely on easily cultured microorganisms such as soil bacteria. When it became apparent that the number of newly discovered compounds could not keep pace with the emerging antibiotic resistances, research expanded towards chemical methods.[10] In combinatorial chemistry, large libraries were generated and screened for bioactivity. Interestingly, chemical properties were found to be distinct between natural products, randomly produced compounds, and drugs.[11] This indicates that random library approaches are not efficient and that natural product-like modifications would be more promising.[8, 10a, 11-12] Nowadays, "diversity-oriented synthesis" is used in the hope to increase the hit rate of synthetic compounds.[10a, 11b, 13] Most of the more recently discovered drugs rely on known active natural product scaffolds that were chemically modified to influence pharmacological properties.[8] In general, the field of natural product research employs a variety of strategies to hunt for novel drug leads. Classical isolation approaches were expanded to include less-studied organisms and those from more obscure origins.[7, 14] Especially marine organisms have been proven to be a rich source of chemically new compounds.[6a, 15] With the advances in diving equipment and robotics, sampling trips are now much more feasible and deeper or more secluded regions can be accessed.[7, 14] Even though many organisms isolated from such more exotic sources are uncultured, recent advances of low-cost sequencing technologies allow us to access this microbial dark matter in an indirect way.[16] Uncultured microorganisms from extreme environments such as Antarctic sponges or tunicates as well as symbionts can now be accessed via metagenome sequencing.[17] Subsequently, genome mining can be applied to identify biosynthetic gene clusters (BGCs) responsible for natural product biosynthesis.[16-18] Such BGCs are found in an incredible range of organisms and suggests that many natural products are yet to be discovered.[16] In addition to classical isolation approaches and (semi-)synthetic chemical approaches, genome mining has turned into a third pillar for drug discovery.[16, 19]

In the following sections, polyketides as a major class of bacterial natural products and their biosynthetic machinery will be introduced. Then, ways to predict polyketide structures will be discussed, followed by ways to biosynthetically diversify polyketides produced by bacteria. Finally, we will dive into genetic recombinational engineering (recombineering) as an option to obtain new natural products.

1.1.2 The discovery of polyketides

One important class of natural products comprises polyketides, named as such by John Norman Collie more than 100 years ago due to the presence of multiple ketone groups.[20] Not only did he coin an entire class of compounds, but he also correctly proposed their mode of assembly via the condensation of acetyl groups.[21] This idea later got more attention as the "polyacetate hypothesis" by Birch and Donovan in the 1950s after acetyl-coenzyme A (acetyl-CoA) was discovered as the biologically active form of acetic acid.[20b, 22] In studies to uncover the biochemistry underlying polyketide biosynthesis, it was shown that the poly-β-keto thioesters formed by head-to-tail linkages of the acetate units could undergo ring closure via aldol-type cyclization or acylation to yield compounds such as 6-methylsalicylic acid (Figure 2).[22a, 23] The "polyacetate hypothesis" was ultimately verified by isotope tracing experiments using labelled acetate and glucose, an upcoming technique at that time.[24]
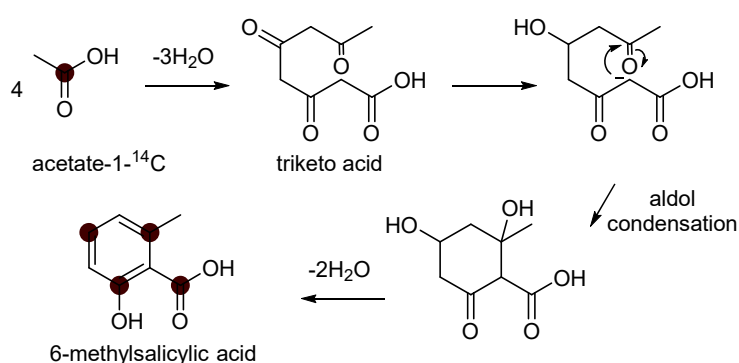

Figure 2. The polyacetate hypothesis as proposed by Birch and Donavan.[22a] Four acetate units can be linked to form a triketo acid. Reduction of one of the keto groups to a hydroxyl group along with carbanion formation then allows for aldol condensation. Different dehydration and enolization reactions finally generate an aromatic product such as 6-methylsalicylic acid. Isotope tracing studies confirmed the incorporation of [14]C (highlighted in black) from acetate-1-[14]C. Figure adapted from Staunton and Weissman[20c].

Only a couple of years later, it was suggested that malonyl-CoA or methylmalonyl-CoA were the actual common extension units in polyketide biosynthesis.[25] In contrast to aromatic polyketides which derive from non-reduced poly-β-keto-chains, the same functionalities can be further reduced or dehydrated in complex polyketides. This feature is reminiscent of fatty acid biosynthesis, a mechanism highly analogous to that of polyketide biosynthesis.

1.2 Fatty acid biosynthesis

Due to its essential function in primary metabolism, fatty acid biosynthesis was studied comprehensively as soon as the first genes were discovered in the 1960s.[26] In fatty acid biosynthesis, a set of enzyme components with distinct functions work together to assemble a fatty acid from thioester-bound acyl units (Figure 3).[20c] First, a 4'-phosphopantetheine (Ppant) transferase (PPTase) attaches the Ppant arm of coenzyme A (CoA) to a conserved serine residue of an acyl carrier protein (ACP).[27] This post-translational modification transforms the apo-ACP into the active holo-ACP which displays a free thiol group for the attachment of intermediates or extender units. Then, a malonyl-acetyl transferase (MAT) loads a starting acyl-CoA unit onto the active site cysteine residue of a ketosynthase (KS). An extending malonate unit is subsequently transferred from its CoA onto the Ppant arm of an ACP. After condensation by the KS, the ACP then delivers the nascent polyketide chain to the active sites in the different enzymes found in the fatty acid synthase (FAS) for adequate processing. The β-keto moiety is sequentially reduced by a ketoreductase (KR), a dehydratase (DH) and an enoyl reductase (ER). After one round of reduction, the growing polyketide chain is transferred back onto the KS and a new extension unit is introduced by the ACP. This reductive cycle is repeated until the fatty acid has a distinct length, after which it is transferred to a thioesterase (TE) responsible for the release of the fatty acid from the FAS.
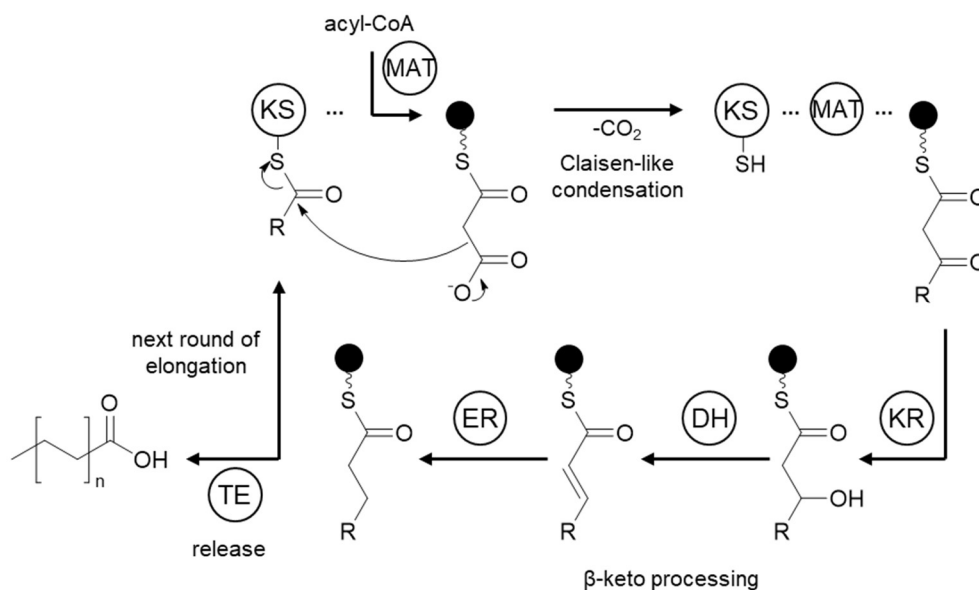

Figure 3. The mechanism of fatty acid biosynthesis. Acyl-CoA building blocks are transferred onto acyl carrier proteins (ACPs, black circles) by malonyl-acetyl transferases (MATs). A ketosynthase (KS) performs condensation and a ketoreductase (KR), a dehydratase (DH), and an enoyl reductase (ER) successively reduce the β-keto moiety to a fully saturated chain. Finally, the fatty acid is released by a thioesterase (TE). Wavy lines symbolize Ppant arms, R: nascent fatty acid chain. Figure adapted from Hertweck [28].

1.3   Polyketide biosynthesis

In general, polyketide biosynthesis follows the same principles and uses the same types of enzymatic transformations as FAS. Aromatic polyketides like actinorhodin or daunorubicin are produced by polyketide synthases (PKSs) with a single set of iteratively acting domains (type II PKSs and iterative type I PKSs).[29] Iterative type III PKSs are independent of ACP-bound substrates and accept CoA-tethered building blocks.[29-30] In this thesis, we focus on the class of non-iterative type I PKSs. These are large multimodular enzymes which rely on multiple sets of domains organized as modules which are each responsible for a single elongation step.[20c, 31] In contrast to FASs, the composition of a module does not necessarily harbor the complete reductive loop (KR, DH, ER). Therefore, the domains present in each module determine the degree of reduction and dehydration. This correlation between polyketide structure and module architecture is commonly referred to as "collinearity rule". The most prominent example of a type I PKS is the deoxyerythronolide B synthase (DEBS) for the production of erythromycin in *Streptomyces* sp. (Figure 4).[32]
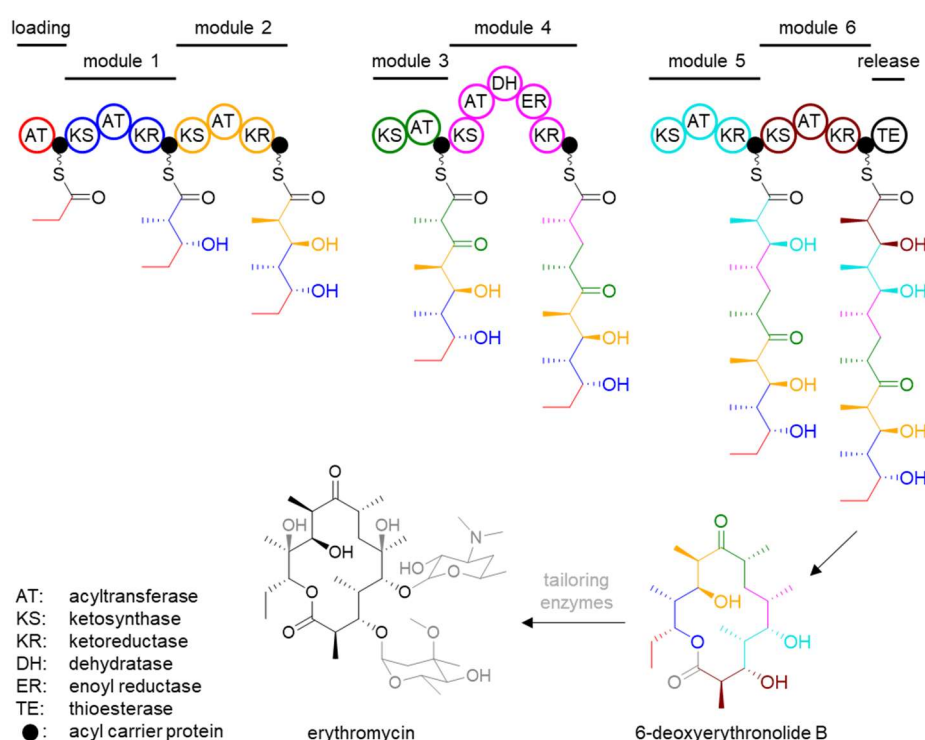

Figure 4. Erythromycin biosynthesis. The colors of the structural moieties correspond to the colors of the module responsible for its installation. Wavy lines indicate Ppant arms. Figure adapted from Staunton and Weissman[20c].

An acyl moiety derived from acyl-CoA is bound to the Ppant arm of a holo-ACP by an acyltransferase (AT) via transesterification. Thioester-bound intermediates are then channeled through the PKS for polyketide formation. A KS catalyzes elongation of the PKS-bound growing polyketide chain via Claisen-like decarboxylative condensation reactions analogous to Figure 3. The resulting β-ketothioester can then be further modified by optional domains. Analogous to their function in fatty acid biosynthesis, KRs, DHs, and ERs reduce the β-keto group to a β-hydroxyl, an α,β double bond, or a fully saturated moiety successively. In addition, methyltransferase (MT) domains can methylate the α-position of the nascent polyketide chain. Along with the incorporation of methylmalonyl building blocks, MTs represent a second way to form α-methylated polyketides. Finally, the thioester-bound intermediate is released by a TE, typically through hydrolysis or macrolactonization. Given that the shuttling of intermediates between modules is reminiscent of the division of labor in Ford's first car manufacturing production lines, PKSs can also be imagined as assembly lines.[33]

16

Another type of megaenzymes relying on modular elongation of a thio-templated intermediate are non-ribosomal peptide synthetases (NRPSs).[34] In contrast to PKSs and FASs, amino acids are used as building blocks in place of acyl units resulting in the production of specific peptide products independent of the ribosome. In a typical NRPS module, an amino acid is activated by an adenylation (A) domain and subsequently thioesterified by a peptidyl carrier protein (PCP). Finally, a condensation (C) domain elongates the growing peptide chain by the introduced amino acid moiety via peptide bond formation.[34b] Given the shared biosynthetic logic of PKSs and NRPSs, it is perhaps unsurprising that the two are compatible and can form hybrid PKS-NRPS systems that produce polyketide-peptide products.

1.4 Structural biology of fatty acid synthases and polyketide synthases

Because of their imperative function in primary metabolism, enzyme structural studies focused mainly on FASs. The high functional similarity between PKSs and metazoan FASs (mFASs) suggests a common evolutionary origin and thus comparable structural architectures.[35] Early biochemical and knock-out experiments indicated that mFASs act as a dimer wherein all enzymatic domains of a subunit closely interact with a central ACP.[36] Consequently, a coiled subunit model was proposed to account for the high interconnectivity required for all domains to interact with the single ACP (Figure 5A).[37]

Motivated by the results obtained for mFAS, researchers focusing on multimodular type I PKSs (hereafter PKSs) performed similar experiments in PKSs, mostly in studies on DEBS. Given that PKS intermediates are channeled through several modules, a certain directionality has to be ensured. In the style of the coiled subunit model for mFAS, a double helical model was postulated for type I PKSs. This so-called "Cambridge model" by Staunton and Leadlay[38] (Figure 5B) was based on a number of limited proteolysis experiments, which indicated that TEs form dimers and that module pairs interact.[38] They speculated that the center of the helix would consist of KS, AT, and ACP domains with any additional domains twisting around in a loop-like fashion. This is in accordance with findings from mutant complementation experiments by Khosla and colleagues.[20c]
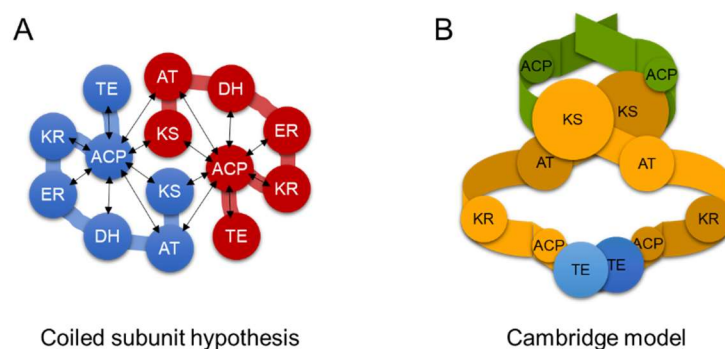

Figure 5. Structural models for mFASs and multimodular non-iterative type I PKSs. A) The coiled subunit hypothesis for FAS dimers. A central ACP interacts with all the enzymatic domains within its subunit. Additionally, the AT, KS, and ACP of each subunit interacts with its counterparts in the other subunit. B) The "Cambridge model" of PKSs as postulated by Staunton and Leadlay. Figure adapted from Staunton[35].

More recently, advances in cryoelectron microscopy (cryo-EM) and X-ray crystallography have provided even greater insights into FAS and PKS structural biology. Most notably, Timm Mayer et al. were able to crystallize an entire mammalian FAS.[39] Even though the overall structure obtained was not resolved at a resolution high enough to identify amino acid side chains, they were able to assign functional domains to electron densities by fitting high-resolution structures from separately crystallized

homologous proteins. In the case of the ACP, its position could only be provisionally assigned to blurred electron densities. This indicated an inherently high mobility, consistent with the hypothesis that the ACP interacts flexibly with the domains within a module. Ban and colleagues further deduced that the active sites of either set of enzymes are facing outwards. Interestingly, the model provided asymmetric reaction chambers, which was explained by the dedication of either chamber to either chain elongation or β-carbon processing. It was further shown that the ACPs can interact with KS and AT of either subunit.[39-40]

during transacylation   during condensation
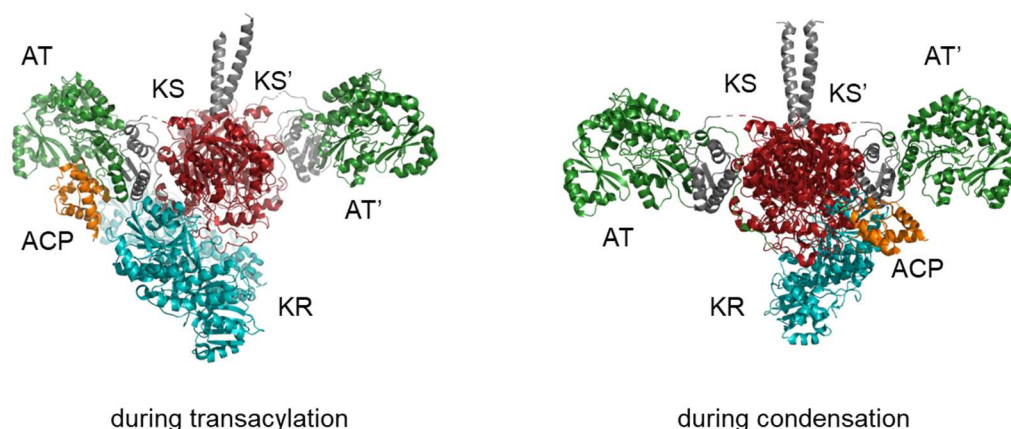
Figure 6. Structures of module 14 from the lasalocid PKS trapped in different conformational states. During transacylation (PDB: 7S6B), the ACP associates with the AT and during condensation (PDB: 7S6C) an ACP-KS interaction can be observed. Figure adapted from Bagde, et al. [41]

The first cryo-EM results for multimodular PKSs unveiled a number of surprises compared to the mFAS structure. The overall architecture seems less relaxed. KSs harbor long helical docking domains at their N-termini which are likely for interaction with the upstream module and some inter-domain linker regions are shorter than in their FAS counterparts.[42] With these linker regions, ACPs can recognize core KS-ATs, but also other protein-protein interfaces.[43] Like in FASs, it is likely that ACPs shuttle the Ppant-tethered intermediate through the PKS by means of random collision. The intermediates bound can induce a conformational change that ensures that the ACP associates with the next domain in line.[44] Such a model would explain the increased reaction time in PKSs compared to FASs.[35, 44a] This was also observed in case of the polyether-producing PKS for lasalocid, where a module was trapped in two distinct states. In the transacylation step, the ACP was found docking to the AT while it was associated with the KS during the condensation step (Figure 6).[41] This indicates that only one chamber can perform polyketide chain extension at the time, in line with findings for FAS.[41, 45] Consistent with this assumption, only one ACP is found docked to another domain in the structures obtained, indicating that the second ACP is not docked at that specific point in time. This "turnstile mechanism" for PKSs was also validated by cryo-EM efforts.[43, 45] According to this newest mechanistic proposal, the KS adopts a closed state until the ACP-tethered intermediate is completely modified by the upstream domains. Only when the ACP transfers the nascent polyketide chain to the downstream module, the KS attains its open conformation again and a new acyl unit is accepted.[43] Two different access points to the KS help ensure the directional biosynthesis in type I PKSs.[46] This way, proper and complete modification is ensured within the individual modules of the assembly line and backwards translocation of the chain is prevented. Concerning the release of the fully elongated polyketide chain, TE domains in multimodular PKSs tend to dimerize while the catalytically active FAS TE remains monomeric. In FAS TEs, the active site consists of a large hydrophobic pocket which is highly selective for reduced fatty acid intermediates of a specific length. In contrast, PKS TEs have a tunnel-like structure which is more flexible considering the length of the substrate.[35] Interestingly, the pH can alter the appearance of the PKS TE substrate channel and is suggested to play a role in hydrolytic versus lactonization release. In case of the cyclizing

DEBS TE, it is proposed, that the bound polyketide chain is shielded from external water that would lead to hydrolysis. Instead, the chain is oriented in a highly specific way that promotes cyclization.[47] All in all, these structural insights provide valuable information in the quest to rewire modules from different PKSs.

1.5 Exploiting the modular architecture of PKSs to produce novel compounds

The modular architecture and the direct relationship between domain order and polyketide structural features make multimodular type I PKSs an interesting target for engineering efforts. Altered assembly lines could be used for the economic production of novel, designed pharmaceutically relevant products inside bacteria. In early experiments, specific domains within a module were inactivated and the expected modified products were observed, further corroborating the collinearity rule.[32] In theory, targeted manipulation of modules and their domains should directly result in the production of specifically modified polyketides. After early type II PKS hybrids between actinorhodin and medermycin or granaticin PKSs were achieved by genetic engineering,[49] initial type I PKS engineering attempts largely focused on DEBS.[20c, 48, 50] Du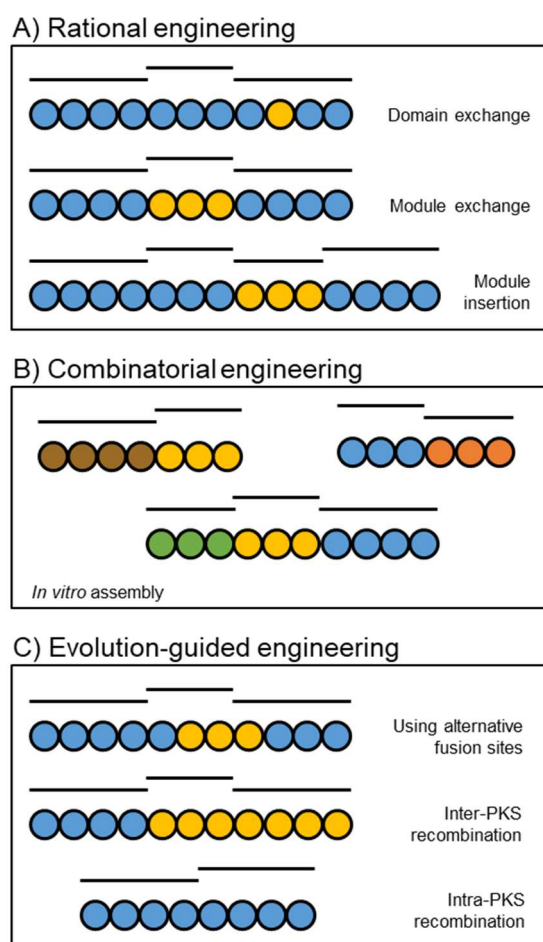e to its relatively compact architecture consisting of just six elongating modules, it turned out to be an excellent model system to study and manipulate PKSs.[20c] The first rationally designed artificial assembly lines relied on the exchange of domains and single modules[51] or the introduction of additional domains into modules[52] (Figure 7A).[53] Some examples include AT domain[51b, 51c] or TE domain swaps.[51d] With advances in *in vitro* methodologies, combinatorial assemblies between different PKS modules were attempted (Figure 7B).[54] However, yields were often way below wild-type production levels, indicating a need for defined protein-protein interactions between native and non-native PKS parts.[48] Structural biology on PKSs and a number of phylogenetic studies have provided key insights into the interconnectivity of PKS domains and modules. This led to a recent shift to more evolution-inspired engineering approaches using alternative fusion sites within conserved inter-domain regions. Additionally, longer PKS parts were introduced leading to shortened or chimeric polyketides instead of products with small alterations from a substituted module or domain (Figure 7C).[48, 55] These identified a natural fusion point at the interface between KS and AT domains that might serve as a point for recombination and rational engineering. The KS's function in accepting substrates from the upstream domains is likely instrumental for the propagation of non-native intermediates, but further insights into the enzymatic interplay between the different PKS

Figure 7. Different engineering strategies employed. Black bars indicate module boundaries. Different colors represent parts from different PKSs. Figure adapted from [48].

parts remains crucial for the controlled design of artificial PKSs.[48, 56] In NRPSs, which share a similar modular biosynthetic logic, chimeric assembly lines were achieved by fusing different modules at short splice points found between C and A domains.[57] Using an exchange unit defined as A-PCP-C, a number of novel peptides have been constructed, some of them at yields comparable to wild-type NRPS.[57-58]

Due to their multimodular architecture, type I PKSs are more flexible than FAS in the number and type of modifications that can be introduced. Each module can be programmed to introduce a specific moiety.[35] To accommodate this flexible biosynthesis, enzymes need to adopt more forgiving substrate binding pockets to accept differentially reduced intermediates. At the same time, the domains accepting a substrate that is passed on from an upstream module need to increase in specificity to ensure that all the expected modifications up to that module have been performed. In FASs, the AT is relatively promiscuous and loads whichever acyl substrate is most abundant.[59] In contrast, type I PKS ATs are highly specific, as exemplified by DEBS ATs in modules 1-6 which only accept methylmalonyl building blocks and thereby act as gatekeepers to determine what type of chain extension occurs.[35] However, the specificity of ATs *can* be engineered to incorporate non-native extender units.[60] Apart from module and substrate pool compositions, other factors readily influence backbone derivatization. Some loading domains comprise ATs specific to uncommon starting building blocks. Similarly, unusual extender units can be integrated, as exemplified by the incorporation of an aminomalonyl building block in zwittermycin[61] or chloroethylmalonate in salinosporamide biosynthesis[62]. As for modifications occurring during the assembly (on-line), internal Diels-Alder reactions, spontaneous heterocyclizations as well as carbacycle formation has been described. Cyclization with and without aromatization are possible tailoring processes. Spontaneous cyclization is common for non-reducing PKSs that generate aromatic polyketides. In contrast to enzymes acting on-line, post-PKS enzymes specifically act on the polyketide once it is released from the PKS. Such enzymes introduce sugar moieties via glycosylation, oxygenations, halogenations or alkylations.[28] Even though post-PKS enzymes add to the diversity observed in polyketides, their high specificity toward a complex polyketide makes them less suitable for engineering purposes.

1.6   *Trans*-acyltransferase polyketide synthases

Compared to canonical type I PKSs like the erythromycin PKS described above, another class of multimodular PKSs was discovered more recently. When the biosynthetic genes (often referred to as biosynthetic gene cluster (BGC) as they are co-localized in the genome) of the pederin,[63] bacillaene,[64] or leinamycin-type[65] polyketides were analyzed, it was found that these lack module-integrated AT domains. Instead, one or more ATs are encoded independently of the core PKS genes. It was proposed to term this class of type I PKSs *trans*-AT PKS and to refer to the textbook multimodular systems as *cis*-AT PKSs.[66] Another name for *trans*-AT systems that is sometimes used is "AT-less".[65] In *trans*-AT PKSs, the minimal module consists of just an ACP and a KS. Interestingly, many of the *trans*-AT PKS-derived polyketides described to date are produced by bacteria closely associated with different organisms.[18b] Some examples include beetles (pederin[18c, 63]), ants (formicolides[67]), sponges (onnamide[18d], peloruside[68], mycalamides[68], pateamides[68], psymberin[69]), tunicates (patellazole[70], mandelalide[71]), plants (gynuellalide[72], lobatamide[72], lacunalide[72], oocydins[73]), lichen (nosperin[74]), bryozoans (bryostatin[75]), fungi (rhizoxin[76]), and even fungi which themselves are associated with spiders (necroxime[77]). Further, a compound was linked to a *trans*-AT PKS in a parasitic protist.[78] Many further complex polyketides have been isolated from marine invertebrates and it is likely that a number of these are in reality produced by a bacterial endosymbiont.

1.6.1 The expanded biosynthetic repertoire of *trans*-AT PKS

Apart from the *trans*-acting ATs, *trans*-AT PKSs also display a number of further peculiarities. In contrast to *cis*-AT PKSs, the biochemistry of *trans*-AT PKSs is not limited to FAS-type domains but includes an expanded range of enzymatic domains.[79] Examples include GCN5-related *N*-acetyl transferase (GNAT) or fatty acyl-AMP ligase (FAAL) loading domains.[27] Further, non-elongating KS domains (KS^{0}s) are found in many *trans*-AT PKSs. The transformation of a β-keto group into a carbon branch (β-branch) by a dedicated enzymatic cassette provides another structural feature observed in many *trans*-AT PKS-derived polyketides.[27] The greater diversity of chemistries installed in *trans*-AT PKS products relative to their *cis*-AT counterparts is exemplified in the biosynthesis of oocydin from *Serratia plymuthica* 4Rx13 (Figure 8). In the first steps, a bisphosphoglycerate loading unit is incorporated which is consecutively transformed into a diene moiety by a β-branching cassette (Figure 8). In the third module, a *trans*-acting Baeyer-Villiger monooxygenase introduces an oxygen atom into the backbone of the growing polyketide chain (Figure 8, marked in red). Hydrolysis of the ester is commonly observed as congener oocydin A and has prolonged assignment of the biosynthetic model due to seemingly superfluous modules upstream of the monooxygenase. Module 5 exhibits yet another unusual module architecture. A second *trans*-acting oxygenase introduces an α-hydroxy group (Figure 8, marked in light blue), a modification orthogonal to standard β-position processing. After acceptance by a specific KS^0, the PKS harbors one of the few canonical PKS modules, followed by a pyran-forming module. Importantly, this module specifically uses the hydroxyl group in the α-position for cyclization. After another set of canonical enzymes, an unprecedented chlorination module follows (Figure 8, marked in green). Concerted action of a *trans*-acting ACP, an Fe(II)/α-ketoglutarate-dependent halogenase, and an auxiliary protein leads to the introduction of a vinyl chloride in the β-γ position. The nascent polyketide chain is further elongated until an internal TE in the penultimate module acetylates the β-hydroxy functionality (Figure 8, marked in orange). Finally, the polyketide is released by macrolactonization (Figure 8, marked in violet). Based on the astonishing density of unusual features, we recently termed it a Swiss army knife PKS.[80]
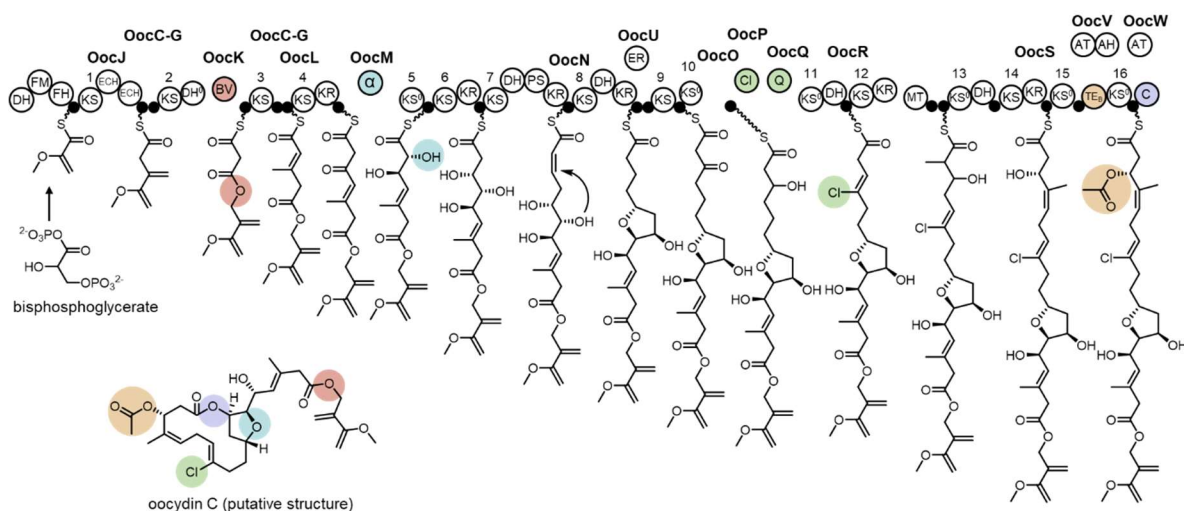

Figure 8. The biosynthetic model of the *trans*-AT PKS for oocydins. DH: dehydratase, FM: methyltransferase, FH: 3-phosphoglycerate-incorporating domain, KS: ketosynthase, ECH: enoyl-CoA hydratase, DH^0: DH-like domain, BV: Baeyer-Villiger monooxygenase, KR: ketoreductase, α: α-hydroxylase, PS: pyran synthase, ER: enoyl reductase, KS^0: non-elongating KS, Cl: Fe(II)/α-ketoglutarate-dependent halogenase, Q: auxiliary protein, MT: methyltransferase, TE$_B$: branching thioesterase, C: condensation domain, AT: acyltransferase, AH: acylhydrolase. Numbers above KS domains refer to the position in the PKS, black circles symbolize acyl carrier proteins, and wavy lines represent Ppant arms. The chemical moieties and the corresponding functionally characterized domains have identical colors. Figure adapted from Hemmerling, et al. [80].

1.6.2   *Trans*-AT PKSs evolution and structural prediction

The presence of *trans*-acting, uncharacterized, or inactive domains greatly complicates prediction of the final polyketide product structure. Further, the extreme diversity observed in *trans*-AT modules means that the KSs of the downstream modules have to evolve specificity to ensure that only completely modified intermediates are channeled through to the next module. This means that the KS might be defined as the end of a functional module. Following that rationale, early phylogenetic analyses of KSs from *trans*-AT PKSs provided first insights into a mode of evolution that seems completely different from what is known for *cis*-AT PKSs.[81] In *cis*-AT PKSs, KSs from one biosynthetic gene cluster usually form distinct clades in a phylogenetic tree, indicating independent and predominantly vertical evolution of multimodular systems.[81] In contrast, KSs in *trans*-AT PKSs form phylogenetic clades according to the chemical structure in the α- to β-position of the incoming intermediates, suggesting extensive horizontal gene transfer as primary mode of evolution.[81] This observation was coined the *trans*-AT PKS correlation rule and consequently applied to KSs of uncharacterized *trans*-AT PKSs to predict the structure of the final product. The predictive power of KSs from *trans*-AT PKSs has been computerized into an algorithm in the tools transATor[79] and transPACT[82], which predict PKS products when provided with a BGC query. Using this tool, a number of novel polyketides derived from *trans*-AT PKSs could be linked to their cognate biosynthetic gene clusters.[79, 81-82] A more recently published bioinformatics tool for the automatic reconstitution of PKS pathways is the Biosynthetic Gene cluster Metabolic pathway Construction (BiGMeC) pipeline.[83]
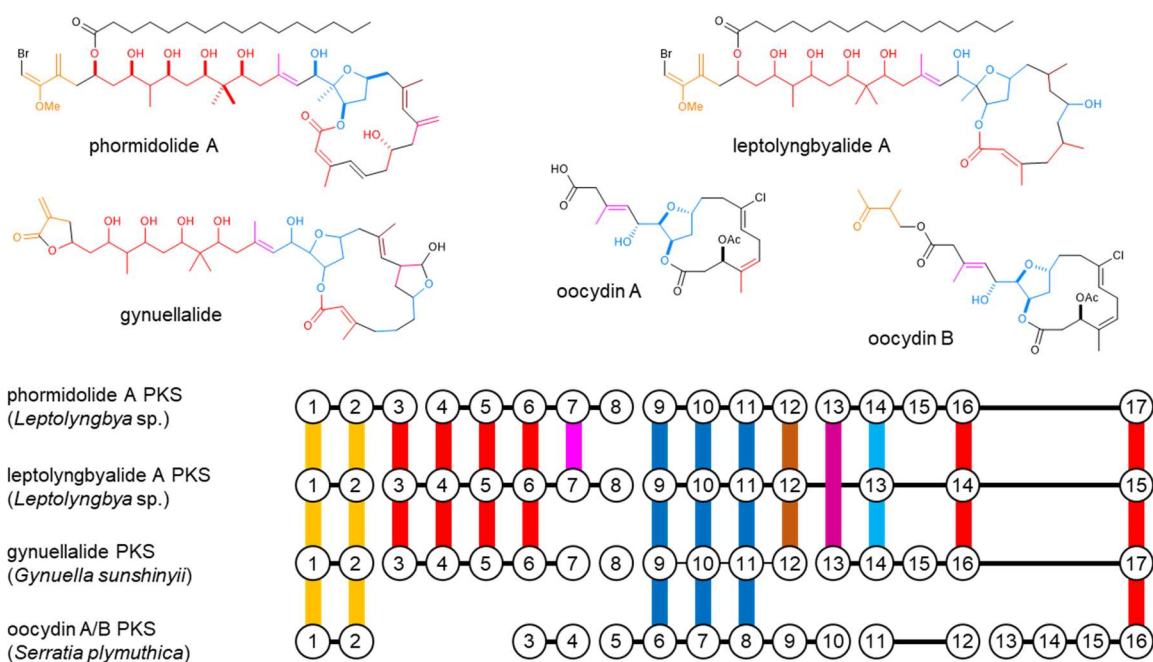

Figure 9. An example of module block evolution. The colors in the chemical structures indicate which modules are responsible for its incorporation. Circles represent KS domains and numbers refer to the position of the KS within the PKS. Figure adapted from Helfrich, et al. [82]

Functional module blocks are found shuffled between clusters in a wide range of organisms.[84] In a recent study, several mosaic *trans*-AT PKSs were identified using the *trans*-AT PKS Annotation and Comparison Tool *trans*PACT, which uses phylogenetic analyses to reveal series of exchanged PKS modules.[82] Interestingly, the exchange of stretches of modules was found in a range of clusters and hints at an increased tolerance of *trans*-AT PKSs compared to *cis*-AT PKSs toward natural recombination. One example of such an extended module swap in the context of oocydin-type polyketides is provided in Figure 9. The phormidolide, leptolyngbyalide, gynuellalide, and oocydin PKSs share a number of modules. While modules 3-6 are conserved in the first three PKSs, the oocydin PKS harbors other, distinct modules. Similarly, the modules between the second and third block of conserved modules differ. The conserved module architectures are also reflected in the final polyketide structures, which share structural features such as the diene moiety or the tetrahydrofuran (Figure 9, marked in yellow and blue, respectively). The natural occurrence of *trans*-AT PKS hybrids renders them exciting targets for the production of artificial assembly lines, especially given that first engineering attempts in *cis*-AT PKSs often proved unsuccessful.[82] Despite the limited knowledge about the structural composition of *trans*-AT PKSs and protein-protein interactions, the natural occurrence of hybrid clusters may indicate a higher plasticity of these megaenzymatic complexes.[48]

1.6.3   Recombineering *trans*-AT PKSs?

The modular fashion and the mosaic-like evolution of *trans*-AT PKSs sparked our interest in generating hybrid artificial assembly lines in these less-studied systems. However, although structural information on *cis*-AT PKS is being produced at an incredible pace, structural insights on *trans*-AT PKSs remain limited to a handful of characterized inter- and intra-domain interactions.[85] Furthermore, modules are often not contained within a single protein, but are split across distinct PKSs and protein boundaries between different domains are therefore common.[27] This is reflected in the identification of intersubunit docking sites that are specific to *trans*-AT PKSs.[86]

1.7   Additional PKS biosynthetic pathways relevant for this thesis

1.7.1   Lobatamide biosynthesis

Lobatamides are benzolactone enamides that were originally isolated from the tunicate *Aplidium lobatum*.[87] They act as V-ATPase inhibitors and feature an unusual oxime moiety. In earlier studies, our lab has postulated a first biosynthetic model for the Lbm *trans*-AT PKS.[88] Apart from a module-integrated oxygenase that performs an integration of an oxygen atom into the polyketide backbone,[88] a second oxygenase is uncharacterized, but predicted to be involved in the installation of a methylated oxime moiety. In Chapter 2, the starting modules of the Lbm *trans*-AT PKS will be studied and linked to the introduction of the oxime moiety and its subsequent methylation.

1.7.2   Bacillaene biosynthesis

The archetypal *trans*-AT PKS is the Bae assembly line responsible for the production of bacillaene in *Bacillus amyloliquefaciens* FZB42.[89] Up to now, several *Bacillus* strains, including the lab strain *Bacillus subtilis* 168 were shown to harbor the *bae* or the related *pks* BGC.[90] Bacillaene is a very unstable polyene compound causing the exact structure to remain elusive for many years until

experiments with liquid chromatography directly coupled to NMR were performed.[91] Functions of bacillaene include antimicrobial activity by inhibiting protein biosynthesis in prokaryotes.[92] In Chapter 3, a first set of rules for recombineering in *trans*-AT PKSs is established in TE exchange experiments using the terminal module of the well-characterized *trans*-AT PKS for bacillaene.

1.8   Approaches to study polyketide synthases

PKSs can be studied on several levels. An indirect way to identify the product of a PKS involves identification of the biosynthetic gene cluster. Advances in genome sequencing along with established bioinformatics tools render genome mining a feasible approach. Among the most used genome mining tools in the context of natural product discovery is antiSMASH.[93] The software includes specific prediction tools for multimodular assembly lines and identifies PKS or NRPS genes. In addition to antiSMASH, the transATor tool was specifically developed for the prediction of *trans*-AT PKS-derived compounds based on the *trans*-AT PKS correlation rule. Such *in silico* studies provide good insights into the architecture of a specific cluster. However, domains can be inactive, act *in trans* or have unknown functions, particularly among *trans*-AT systems. One way to decipher the enzymology performed by these modules is through biochemical studies. Due to their gigantic nature, whole protein expression for *in vitro* studies is currently only practical for truncated assembly lines.[94] An impressive effort was the complete reconstitution of the nocardiosis-associated *cis*-/*trans*-AT hybrid PKS assembly line in *Escherichia coli* (*E. coli*) and the identification of its associated polyketide product.[95] While heterologous expression in hosts like *E. coli* or *Streptomyces* is possible, several factors have to be taken into account. Substrate acyl-CoAs have to be present at sufficient quantities and PPTases for the post-translational activation of *apo*-ACPs are also required. Furthermore, transporters for export or resistance genes for the tolerance of the potentially toxic polyketide products may be needed.[94, 96] A more facile approach is the study of specific domains with unknown function through *in vitro* assays. The domains of interest are heterologously expressed and subsequently assayed with the expected substrate based on the biosynthetic model. Given that mid-assembly line intermediates can be complex, unknown, and/or hard to synthesize, simplified analogs are valuable alternatives. In particular, *N*-acetylcysteamine (NAC)-linked probes are widely used to circumvent the high costs and inherent instability of CoA-substrates.[97] NAC thioesters (SNACs) can pass membranes and can thus even be used in *in vivo* assays. Due to their structural similarity, they can often be used as surrogates for CoA or ACP-bound substrates (Figure 10). If necessary, a holo-ACP heterologously produced in an *E. coli* strain harboring a PPTase can quickly be loaded with the corresponding SNAC.[97] Lastly, SNACs can be modified to crosslink domains, to tag ACPs via click-chemistry or to offload biosynthetic intermediates from the PKS.[98]
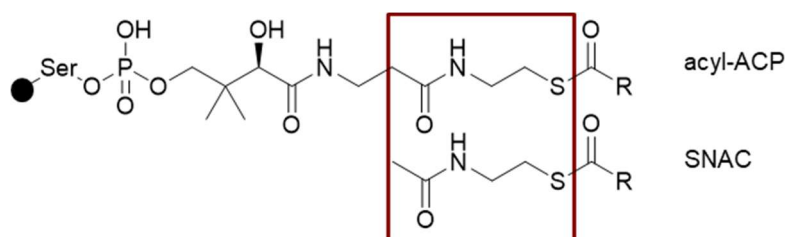

Figure 10. Acyl-ACP and SNAC structures compared. The box indicates the region the SNAC mimics. Black circles represent ACPs and the wavy line a Ppant arm.

1.9   Aims of this thesis

The overarching theme of this thesis is to engineer artificial *trans*-AT PKS assembly lines. Such chimeric assembly lines could be used to produce new drug candidates in an economic way, evading complicated chemical syntheses. In our quest for a plug-and-play setup to produce complex polyketides in a bacterial host, we start gathering stepwise information on *trans*-AT PKSs biosynthesis. To do so, we first set out to biochemically characterize a new set of modules responsible for on-PKS methoximation. The characterization of such new modules is not only important to gain further insight into the enzymatic repertoire of *trans*-AT PKSs, but also plays a key role in the improvement of the predictive power of analysis tools. Furthermore, we aim to define a basic set of rules to allow shuffling of different modules. A central prerequisite for any engineered *trans*-AT PKS is the processing and final release of non-native substrates. We therefore first focus on releasing modules. In studies on a broad range of offloading modules, we establish guidelines for offloading of non-native polyketide products. Finally, we employ a versatile but also highly complex *trans*-AT PKS as a platform to study intermediate channeling between non-native modules. We make use of the intricate interactions between an exotic halogenating protein pair to introduce foreign PKS parts on a plasmid to the native host. Further, we use statistical coupling analysis to pinpoint a potential fusion site between modules and experimentally test it in a series of chimeric *trans*-AT PKSs. With these studies we aim to explore ways to engineer *trans*-AT PKSs to lay the foundation for the bacterial production of novel polyketide natural products.

References

[1]   A. A. Brakhage, *Nat. Rev. Microbiol.* **2013**, *11*, 21-32.

[2]   J. H. Crosa, C. T. Walsh, *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 223-249.

[3]   L. J. Cseke, P. B. Kaufman, A. Kirakosyan, *The Biology of Essential Oils in the Pollination of Flowers, Vol. 2*, **2007**.

[4]   R. L. Kellner, K. Dettner, *Oecologia* **1996**, *107*, 293-300.

[5]   a) J. Clardy, C. Walsh, *Nature* **2004**, *432*, 829-837; b) J. Davies, *J. Antibiot. Res.* **2013**, *66*, 361-364; c) A. M. Sayed, M. H. A. Hassan, H. A. Alhadrami, H. M. Hassan, M. Goodfellow, M. E. Rateb, *J. Appl. Microbiol.* **2020**, *128*, 630-657.

[6]   a) Y.-W. Chin, M. J. Balunas, H. B. Chai, A. D. Kinghorn, *AAPS J.* **2006**, *8*, E239-E253; b) D. J. Newman, G. M. Cragg, K. M. Snader, *Nat. Prod. Rep.* **2000**, *17*, 215-234; c) D. A. Dias, S. Urban, U. Roessner, *Metabolites* **2012**, *2*, 303-336.

[7]   G. M. Cragg, D. J. Newman, *Pure Appl. Chem.* **2005**, *77*, 7-24.

[8]   D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2020**, *83*, 770-803.

[9]   H. Wang, C. Tang, Y. Wang, M. Han, F. Jiang, L. Jiang, J. Wu, C. Fu, Y. Chen, Q. Jiang, *Environ. Pollut.* **2021**, *291*, 118167.

[10]   a) I. Paterson, E. A. Anderson, *Science* **2005**, *310*, 451-453; b) P. M. Wright, I. B. Seiple, A. G. Myers, *Angew. Chem. Int. Ed.* **2014**, *53*, 8840-8869.

[11]   a) T. H. Keller, A. Pichota, Z. Yin, *Curr. Opin. Chem. Biol.* **2006**, *10*, 357-361; b) T. Henkel, R. M. Brunne, H. Müller, F. Reichel, *Angew. Chem. Int. Ed.* **1999**, *38*, 643-647.

[12]   D. J. Newman, G. M. Cragg, K. M. Snader, *J. Nat. Prod.* **2003**, *66*, 1022-1037.

[13]   a) C. A. Lipinski, *Drug Discov. Today Technol.* **2004**, *1*, 337-341; b) M. Feher, J. M. Schmidt, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218-227.

[14]   D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2004**, *67*, 1216-1238.

[15]   D.-X. Kong, Y.-Y. Jiang, H.-Y. Zhang, *Drug Discov. Today* **2010**, *15*, 884-886.

[16]   G. M. Cragg, D. J. Newman, *Biochim. Biophys. Acta Gen. Subj.* **2013**, *1830*, 3670-3695.

[17]   Z. E. Wilson, M. A. Brimble, *Nat. Prod. Rep.* **2009**, *26*, 44-71.

[18]   a) O. K. Radjasa, Y. M. Vaske, G. Navarro, H. C. Vervoort, K. Tenney, R. G. Linington, P. Crews, *Bioorg. Med. Chem.* **2011**, *19*, 6658-6674; b) J. Piel, *Nat. Prod. Rep.* **2009**, *26*, 338-362; c) J. Piel, D. Butzke, N. Fusetani, D. Hui, M. Platzer, G. Wen, S. Matsunaga, *J. Nat. Prod.* **2005**, *68*, 472-479; d) J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16222-16227; e) L. P. Partida-Martinez, C. Hertweck, *Nature* **2005**, *437*, 884-888; f) L. P. Partida-Martinez, C. Flores de Looß, K. Ishida, M. Ishida, M. Roth, K. Buder, C. Hertweck, *Appl. Environ. Microbiol.* **2007**, *73*, 793-797.

[19]   E. S. Gomes, V. Schuch, E. G. d. M. Lemos, *Braz. J. Microbiol.* **2013**, *44*, 1007-1034.

[20]   a) J. N. Collie, *J. Chem. Soc., Trans.* **1907**, *91*, 1806-1813; b) R. Bentley, *Crit. Rev. Biotechnol.* **1999**, *19*, 1-40; c) J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380-416.

[21]   J. N. Collie, *J. Chem. Soc., Trans.* **1893**, *63*, 329-337.

[22]   a) A. Birch, F. Donovan, *Aust. J. Chem.* **1953**, *6*, 360-368; b) A. J. Birch, *Science* **1967**, *156*, 202-206.

[23]   R. Bentley, J. W. Bennett, *Annu. Rev. Microbiol.* **1999**, *53*, 411-446.

[24]   A. Birch, R. Massy-Westropp, C. Moye, *Aust. J. Chem.* **1955**, *8*, 539-544.

[25]   F. Lynen, *J. Cell. Comp. Physiol.* **1959**, *54*, 33-49.

[26]   a) A. J. Fulco, *Prog. Lipid Res.* **1983**, *22*, 133-160; b) R. O. Brady, *J. Biol. Chem.* **1960**, *235*, 3099-3103; c) R. O. Brady, R. M. Bradley, E. G. Trams, *J. Biol. Chem.* **1960**, *235*, 3093-3098; d) S. J. Wakil, *Am. J. Clin. Nutr.* **1960**, *8*, 630-644.

[27]   E. J. N. Helfrich, J. Piel, *Nat. Prod. Rep.* **2016**, *33*, 231-316.

[28]   C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688-4716.

[29]   B. Shen, *Curr. Opin. Chem. Biol.* **2003**, *7*, 285-295.

[30]   K. J. Weissman, in *Methods Enzymol., Vol. 459*, Academic Press, **2009**, pp. 3-16.

[31]   H. Jenke-Kodama, E. Dittmann, *Phytochemistry* **2009**, *70*, 1858-1866.

[32]   S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **1991**, *252*, 675-679.

[33]   N. M. Llewellyn, J. B. Spencer, *Nature* **2007**, *448*, 755-756.

[34] a) L. Du, C. Sánchez, B. Shen, *Metab. Eng.* **2001**, *3*, 78-95; b) R. Finking, M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453-488; c) R. D. Süssmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770-3821.

[35] S. Smith, S.-C. Tsai, *Nat. Prod. Rep.* **2007**, *24*, 1041-1072.

[36] A. Witkowski, A. Joshi, S. Smith, *Biochem.* **1996**, *35*, 10569-10575.

[37] A. Witkowski, A. Ghosal, A. K. Joshi, H. E. Witkowska, F. J. Asturias, S. Smith, *Chem. Biol.* **2004**, *11*, 1667-1676.

[38] J. Staunton, P. Caffrey, J. F. Aparicio, G. A. Roberts, S. S. Bethell, P. F. Leadlay, *Nat. Struct. Biol.* **1996**, *3*, 188-192.

[39] T. Maier, S. Jenni, N. Ban, *Science* **2006**, *311*, 1258-1262.

[40] T. Maier, M. Leibundgut, N. Ban, *Science* **2008**, *321*, 1315-1322.

[41] S. R. Bagde, I. I. Mathews, J. C. Fromme, C.-Y. Kim, *Science* **2021**, *374*, 723-729.

[42] J. L. Smith, G. Skiniotis, D. H. Sherman, *Curr. Opin. Struct. Biol.* **2015**, *31*, 9-19.

[43] B. Lowry, X. Li, T. Robbins, D. E. Cane, C. Khosla, *ACS Cent. Sci.* **2016**, *2*, 14-20.

[44] a) M. Klaus, E. Rossini, A. Linden, K. S. Paithankar, M. Zeug, Z. Ignatova, H. Urlaub, C. Khosla, J. Köfinger, G. Hummer, M. Grininger, *J. Am. Chem. Soc. Au* **2021**; b) J. R. Whicher, S. Dutta, D. A. Hansen, W. A. Hale, J. A. Chemler, A. M. Dosey, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith, G. Skiniotis, *Nature* **2014**, *510*, 560-564.

[45] D. P. Cogan, K. Zhang, X. Li, S. Li, G. D. Pintilie, S.-H. Roh, C. S. Craik, W. Chiu, C. Khosla, *Science* **2021**, *374*, 729-734.

[46] S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith, G. Skiniotis, *Nature* **2014**, *510*, 512-517.

[47] S.-C. Tsai, H. Lu, D. E. Cane, C. Khosla, R. M. Stroud, *Biochem.* **2002**, *41*, 12598-12606.

[48] A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, *Chem. Rev.* **2019**, *119*, 12524-12547.

[49] a) D. A. Hopwood, F. Malpartida, H. Kieser, H. Ikeda, J. Duncan, I. Fujii, B. Rudd, H. Floss, S. Ōmura, *Nature* **1985**, *314*, 642-644; b) H. G. Floss, *Trends Biotechnol.* **1987**, *5*, 111-115.

[50] B. J. Rawlings, *Nat. Prod. Rep.* **2001**, *18*, 190-227.

[51] a) R. McDaniel, C. M. Kao, S. J. Hwang, C. Khosla, *Chem. Biol.* **1997**, *4*, 667-674; b) M. Oliynyk, M. J. Brown, J. Cortés, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1996**, *3*, 833-839; c) X. Ruan, A. Pereda, D. L. Stassi, D. Zeidner, R. G. Summers, M. Jackson, A. Shivakumar, S. Kakavas, M. J. Staver, S. Donadio, *J. Bacteriol.* **1997**, *179*, 6416-6425; d) R. S. Gokhale, D. Hunziker, D. E. Cane, C. Khosla, *Chem. Biol.* **1999**, *6*, 117-125; e) A. Ranganathan, M. Timoney, M. Bycroft, J. Cortés, I. P. Thomas, B. Wilkinson, L. Kellenberger, U. Hanefeld, I. S. Galloway, J. Staunton, *Chem. Biol.* **1999**, *6*, 731-741.

[52] A. F. A. Marsden, B. Wilkinson, J. Cortés, N. J. Dunster, J. Staunton, P. F. Leadlay, *Science* **1998**, *279*, 199-202.

[53] C. J. Rowe, I. U. Böhm, I. P. Thomas, B. Wilkinson, B. A. Rudd, G. Foster, A. P. Blackaby, P. J. Sidebottom, Y. Roddis, A. D. Buss, *Chem. Biol.* **2001**, *8*, 475-485.

[54] a) P. F. Leadlay, *Curr. Opin. Chem. Biol.* **1997**, *1*, 162-168; b) H. G. Menzella, C. D. Reeves, *Curr. Opin. Microbiol.* **2007**, *10*, 238-245; c) H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood, D. V. Santi, *Nat. Biotechnol.* **2005**, *23*, 1171-1176.

[55] a) L. Kellenberger, I. S. Galloway, G. Sauter, G. Böhm, U. Hanefeld, J. Cortés, J. Staunton, P. F. Leadlay, *Chembiochem* **2008**, *9*, 2740-2749; b) Y. Sugimoto, L. Ding, K. Ishida, C. Hertweck, *Angew. Chem. Int. Ed.* **2014**, *53*, 1560-1564; c) Y. Sugimoto, K. Ishida, N. Traitcheva, B. Busch, H.-M. Dahse, C. Hertweck, *Chem. Biol.* **2015**, *22*, 229-240; d) A. Wlodek, S. G. Kendrew, N. J. Coates, A. Hold, J. Pogwizd, S. Rudder, L. S. Sheehan, S. J. Higginbotham, A. E. Stanley-Smith, T. Warneck, *Nat. Commun.* **2017**, *8*, 1-10; e) W. B. Porterfield, N. Poenateetai, W. Zhang, *iScience* **2020**, *23*, 100938.

[56] M. Klaus, M. Grininger, *Nat. Prod. Rep.* **2018**, *35*, 1070-1081.

[57] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275-281.

[58] a) K. A. J. Bozhüyük, A. Linck, A. Tietze, J. Kranz, F. Wesche, S. Nowak, F. Fleischhacker, Y.-N. Shi, P. Grün, H. B. Bode, *Nat. Chem.* **2019**, *11*, 653-661; b) K. Bozhüyük, J. Watzel, N. Abbood, H. B. Bode, *Angew. Chem. Int. Ed.* **2021**; c) H.-M. Huang, P. Stephan, H. Kries, *Cell Chem. Biol.* **2021**, *28*, 221-227.e227.

[59] Y.-K. Park, F. Bordes, F. Letisse, J.-M. Nicaud, *Metab. Eng. Commun.* **2021**, *12*, e00158.

[60] S. M. Carpenter, G. J. Williams, *ACS Chem. Biol.* **2018**, *13*, 3361-3373.

[61] Y. A. Chan, M. T. Boyne, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher, M. G. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14349.

[62] A. S. Eustáquio, R. P. McGlinchey, Y. Liu, C. Hazzard, L. L. Beer, G. Florova, M. M. Alhamadsheh, A. Lechner, A. J. Kale, Y. Kobayashi, K. A. Reynolds, B. S. Moore, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12295.

[63] J. Piel, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14002-14007.

[64] A. M. Albertini, T. Caramori, F. Scoffone, C. Scotti, A. Galizzi, *Microbiology* **1995**, *141*, 299-309.

[65] Y.-Q. Cheng, G.-L. Tang, B. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3149-3154.

[66] J. Piel, G. Wen, M. Platzer, D. Hui, *Chembiochem* **2004**, *5*, 93-98.

[67] J. S. An, J. Y. Lee, E. Kim, H. Ahn, Y.-J. Jang, B. Shin, S. Hwang, J. Shin, Y. J. Yoon, S. K. Lee, D.-C. Oh, *J. Nat. Prod.* **2020**, *83*, 2776-2784.

[68] M. Rust, E. J. Helfrich, M. F. Freeman, P. Nanudorn, C. M. Field, C. Rückert, T. Kündig, M. J. Page, V. L. Webb, J. Kalinowski, *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 9508-9518.

[69] K. M. Fisch, C. Gurgui, N. Heycke, S. A. Van Der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, *Nat. Chem. Biol.* **2009**, *5*, 494.

[70] J. C. Kwan, M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood, E. W. Schmidt, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 20655-20660.

[71] J. Lopera, I. J. Miller, K. L. McPhail, J. C. Kwan, *mSystems* **2017**, *2*, e00096-00017.

[72] R. Ueoka, A. Bhushan, S. I. Probst, W. M. Bray, R. S. Lokey, R. G. Linington, J. Piel, *Angew. Chem. Int. Ed.* **2018**, *130*, 14727-14731.

[73] M. A. Matilla, H. Stöckmann, F. J. Leeper, G. P. Salmond, *J. Biol. Chem.* **2012**, *287*, 39125-39138.

[74] A. Kampa, A. N. Gagunashvili, T. A. Gulder, B. I. Morinaka, C. Daolio, M. Godejohann, V. P. Miao, J. Piel, Ó. S. Andrésson, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3129-E3137.

[75] S. Sudek, N. B. Lopanik, L. E. Waggoner, M. Hildebrand, C. Anderson, H. Liu, A. Patel, D. H. Sherman, M. G. Haygood, *J. Nat. Prod.* **2007**, *70*, 67-74.

[76] L. P. Partida-Martinez, C. Hertweck, *Chembiochem* **2007**, *8*, 41-45.

[77] S. P. Niehs, B. Dose, S. Richter, S. J. Pidot, H. M. Dahse, T. P. Stinear, C. Hertweck, *Angew. Chem. Int. Ed.* **2020**, *59*, 7766-7771.

[78] G. Zhu, M. J. LaGier, F. Stejskal, J. J. Millership, X. Cai, J. S. Keithly, *Gene* **2002**, *298*, 79-89.

[79] E. J. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, *Nat. Chem. Biol.* **2019**, *15*, 813-821.

[80] F. Hemmerling, R. A. Meoded, A. E. Fraley, H. A. Minas, C. L. Dieterich, M. Rust, R. Ueoka, K. Jensen, E. J. Helfrich, C. Bergande, M. Biedermann, N. Magnus, B. Piechulla, J. Piel, *Angew. Chem. Int. Ed.* **2022**.

[81] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225-233.

[82] E. J. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, *Nat. Commun.* **2021**, *12*, 1-14.

[83] S. Sulheim, F. A. Fossheim, A. Wentzel, E. Almaas, *BMC Bioinform.* **2021**, *22*, 81.

[84] R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, *Nat. Chem. Biol.* **2015**, *11*, 705-712.

[85] a) S. Kosol, M. Jenner, J. R. Lewandowski, G. L. Challis, *Nat. Prod. Rep.* **2018**, *35*, 1097-1109; b) M. Passmore, A. Gallo, J. R. Lewandowski, M. Jenner, *Chem. Sci.* **2021**, *12*, 13676-13685.

[86] a) F. Risser, S. Collin, R. Dos Santos-Morais, A. Gruez, B. Chagot, K. J. Weissman, *J. Struct. Biol.* **2020**, *212*, 107581; b) M. Jenner, S. Kosol, D. Griffiths, P. Prasongpholchai, L. Manzi, A. S. Barrow, J. E. Moses, N. J. Oldham, J. R. Lewandowski, G. L. Challis, *Nat. Chem. Biol.* **2018**, *14*, 270-275; c) J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2016**, *11*, 2466-2474; d) J. L. Meinke, A. J. Simon, D. T. Wagner, B. R. Morrow, S. You, A. D. Ellington, A. T. Keatinge-Clay, *ACS Synth. Biol.* **2019**, *8*, 2017-2024.

[87] a) T. C. McKee, D. L. Galinis, L. K. Pannell, J. H. Cardellina, J. Laakso, C. M. Ireland, L. Murray, R. J. Capon, M. R. Boyd, *J. Org. Chem.* **1998**, *63*, 7805-7810; b) D. L. Galinis, T. C. McKee, L. K. Pannell, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **1997**, *62*, 8968-8969.

[88]   R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem. Int. Ed.* **2020**, *59*, 7761-7765.

[89]   a) J. Moldenhauer, X.-H. Chen, R. Borriss, J. Piel, *Angew. Chem. Int. Ed.* **2007**, *46*, 8195-8197; b) X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, G. Gottschalk, R. D. Süssmuth, R. Borriss, *J. Bacteriol.* **2006**, *188*, 4024-4036.

[90]   G. Aleti, A. Sessitsch, G. Brader, *Comput. Struct. Biotechnol. J.* **2015**, *13*, 192-203.

[91]   R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506.

[92]   P. S. Patel, S. HuANG, S. Fisher, D. Pirnik, C. Aklonis, L. Dean, E. Meyers, P. Fernandes, F. Mayerl, *J. Antibiot.* **1995**, *48*, 997-1003.

[93]   K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81-W87.

[94]   Y. Gao, Y. Zhao, X. He, Z. Deng, M. Jiang, *Curr. Opin. Biotechnol.* **2021**, *69*, 103-111.

[95]   K. P. Yuet, C. W. Liu, S. R. Lynch, J. Kuo, W. Michaels, R. B. Lee, A. E. McShane, B. L. Zhong, C. R. Fischer, C. Khosla, *J. Am. Chem. Soc.* **2020**, *142*, 5952-5957.

[96]   A. Pfeifer Blaine, C. Khosla, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 106-118.

[97]   J. Franke, C. Hertweck, *Cell Chem. Biol.* **2016**, *23*, 1179-1192.

[98]   I. Wilkening, S. Gazzola, E. Riva, J. S. Parascandolo, L. Song, M. Tosin, *ChemComm* **2016**, *52*, 10392-10395.

Chapter II

Manuscript in preparation

Modular Oxime Formation by Polyketide Synthases

Hannah A. Minas[1], Franziska Hemmerling[1], Roy A. Meoded[1], Silke I. Probst[1], Cora L. Dieterich[1], Simon Rüdisser[2], and Jörn Piel*[1]

[1]Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zürich, Vladimir-Prelog-Weg 4, 8093 Zürich (Switzerland).
E-mail: jpiel@ethz.ch

[2]Institute of Molecular Biology and Biophysics, Biomolecular NMR Spectroscopy Platform, Eidgenössische Technische Hochschule (ETH) Zürich, Hönggerbergring 64, 8093 Zürich (Switzerland)

Author contributions

HAM, FH, RAM, and JP designed the research. HAM performed cloning, expression experiments, assays, HPLC-MS analysis, and analyzed the data. HAM and FH performed protein purification, HAM, RAM, SIP, and CLD performed synthesis. HAM and SR performed ^{15}N-NMR analysis, HAM and CLD wrote the supporting information. HAM and JP wrote the manuscript with input from all authors.

Abstract

Trans-acyltransferase polyketide synthases (*trans*-AT PKSs) are enzymatic assembly lines responsible for the biosynthesis of complex natural products. They act as a series of modules elongating carrier protein-bound intermediates. A unique feature of *trans*-AT PKSs are ketosynthases (KSs) that phylogenetically clade according to the incoming substrate. This property allows the prediction of enzymology occurring upstream of the KS domain analyzed. Here we use unassigned KS clades to unravel exotic biochemistry in the lobatamide *trans*-AT PKS. *In vitro* and *in vivo* assays support the installation of a methylated oxime group as part of on-line modifications performed by an unusual bimodule. These newly characterized oxime-forming modules showcase the versatility of *trans*-AT PKS biochemistry and improve the predictive power of KS phylogeny as a tool to decipher these megaenzymes.

Introduction

Benzolactone enamides are a family of cytotoxic vacuolar-type ATPase (V-ATPase)-inhibiting polyketide natural products[1] reported from a broad range of organisms including fungi,[2] sponges,[3] tunicates,[4] and bacteria[5] (Figure 1). Many representatives contain, besides the benzolactone and enamide functions, an unusual *O*-methyl oxime moiety that is present in few other known polyketides.[6] Recently, first insights into the biosynthetic origin of the oxime-bearing lobatamide A (**1**) and necroxime A (**2**) were reported for two bacterial producers, the plant root-associated *Gynuella sunshinyii* YC6258[7] and the fungal endosymbiont *Burkholderia* sp. HKI-0404.[8] Their host-associated lifestyles and the diverse reported natural product sources suggest that V-ATPase inhibitors mediate symbiotic interactions in a broad range of organisms.

Lobatamides and necroximes are generated by highly complex multimodular biosynthetic enzymes termed *trans*-acyltransferase polyketide synthases (*trans*-AT PKSs). In general, PKS modules sequentially elongate and functionalize an acyl carrier protein (ACP)-bound growing polyketide chain in a stepwise fashion. In each module, a ketosynthase (KS) domain elongates the growing polyketide chain by a ketide unit supplied by an acyltransferase (AT).[9] The introduced building block can be further processed by optional domains including ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs) to successively reduce the β-keto group to β-hydroxy, α,β-double-bond, or fully saturated moieties.[10] PKSs can also include modules from non-ribosomal peptide synthetases (NRPS) that incorporate amino acids into the growing chain via the concerted action of adenylation (A) and condensation domains (C).[9a, 11] An intriguing feature of *trans*-AT PKSs is the abundance of non-canonical modules catalyzing various additional reactions.[10, 12]. In these assembly lines, a perplexing array of module-integrated or lone standing enzymes can act during chain elongation. To facilitate genome mining and biosynthetic studies for these challenging enzymes, our group has previously developed a predictive method,[13] implemented in the computational tools TransATor[14] and transPACT,[12a] that infers polyketide moieties introduced by *trans*-AT PKS modules from the sequence of coevolving downstream KS domains. Thus, the polyketide unit introduced by a module correlates with the phylogenetic clade of the KS in the downstream module. This method greatly streamlines the prediction and discovery of *trans*-AT PKS products but reaches its limit when novel module functions are encountered that result in unassignable KS clades.[12a] Here we report biochemical studies on the V-ATPase inhibitor lobatamide A[15] (**1**, Figure 1A) that demonstrate oxime installation as a novel, integral function of *trans*-AT PKSs. The modification is accomplished by a trimodular system with oxygenating and methylating components that is conserved in all known benzolactone oxime pathways (Figures 1A-C). The results expand the biosynthetic scope of *trans*-AT assembly lines and offer potential to introduce oxime tags into engineered complex polyketides.

Results and Discussion

Based on the PKS architecture and the structure of the mature benzolactone polyketides, we and others[7-8, 16] hypothesized that the methylated oxime moiety is not installed by post-PKS modifications but by the first three modules of the assembly line. These modules are also found in the PKS-NRPSs of two oximidine III (**3**) producers (Figure 1C). The biosynthetic assignment is further supported by the absence of the modules in the *trans*-AT PKS of the related compound apicularen A (**4**) lacking the oxime unit (Figure 1D).[8] In the lobatamide pathway, the first module has the hallmarks of an NRPS loading module with an A domain and a peptidyl carrier protein (PCP) domain[8, 17]. Analysis of the A domain by NRPSpredictor2[18] of antiSMASH[19] identifies a sequence signature suggesting that glycine is loaded onto the downstream PCP.
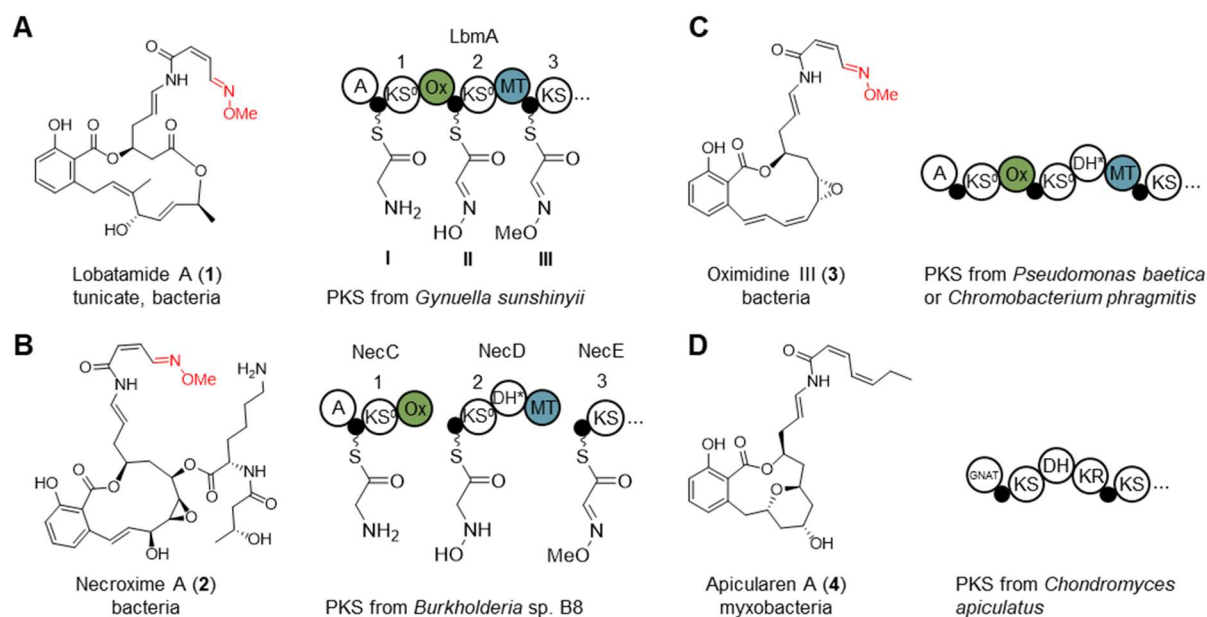
Figure 1. Selected examples of *trans*-AT PKS-derived benzolactone enamide polyketide natural products and associated biosynthetic gene clusters. If present, the oxime group is marked in red. The compound source is given beneath the name. A) Proposed first steps of lobatamide A biosynthesis with the respective portion of the PKS protein LbmA shown. The consecutive action of LbmA-Ox and LbmA-MT is hypothesized to install an oxime moiety which is subsequently transformed into a methylated oxime moiety. B) Domain representation of the *trans*-AT PKS for **2** on the respective PKS proteins NecC, NecD, NecE, *indicates that the domain is absent in the PKS from the alternative necroxime producer *Rhizobium* sp. BK314. C) Domain representation of the *trans*-AT PKS for **3**. *indicates that the domain is absent in the PKS from *Pseudomonas baetica*. D) Domain representation of the *trans*-AT PKS for **4**. A: adenylation domain, KS: ketosynthase, KS^0: non-elongating ketosynthase, Ox: Flavin-dependent monooxygenase, MT: methyl transferase, KR: ketoreductase, DH: dehydratase, GCN5-related *N*-acetyl transferase, black circles: carrier protein domains. The numbers above the KS indicate the KS's position in the PKS. Roman numbers refer to the putative thioester-bound intermediates.

Congruously, the following module harbors a non-elongating KS^0 domain (LbmA-$KS^0$1) that phylogenetically clades with glycine-accepting KSs.[7] This PKS module further harbors a non-canonical domain (LbmA-Ox) with homology to flavin-dependent monooxygenases as a candidate for hydroxylation of the thioester-linked intermediate **I** (Figure 1A), which we hypothesized would be subsequently *O*-methylated by the methyltransferase (MT) of the third module (LbmA-MT). The related *trans*-AT PKS assembly lines for necroxime and oximidine each harbor an additional DH domain that was postulated to assist in conversion of a hydroxylated glycine to an oxime.[8] In our previous analyses, an automated prediction of the substrates for LbmA-$KS^0$2 and LbmA-KS3 using TransATor[14] did not provide conclusive structural suggestions.[7] We therefore inferred an improved phylogenetic tree using a collection of more than 1000 annotated KS sequences from *trans*-AT PKSs obtained from NCBI and found LbmA-$KS^0$2 and LbmA-KS3 in unassigned clades that also contained the manually added KSs from the corresponding modules in the necroxime and oximidine *trans*-AT PKS systems (Figure S1). This prompted us to investigate the enzymology happening in the respective upstream modules to characterize their potentially new function and to assign the KSs. To test the role of the putative oxygenase LbmA-Ox in oxime formation, we cloned the corresponding region of the *lbm* biosynthetic gene cluster into a derivative of the pET28b expression vector (pET28b-SUMO, provided by Prof. Groll). Upon expression in *E. coli* as an N-terminally His₆- and SUMO-tagged protein (Figure S2) and purification, yellow protein was obtained indicating bound flavin adenine dinucleotide (FAD). As surrogate for intermediates tethered via 4'-phosphopantetheinyl moieties to the PKS, *N*-

acetylcysteamine (SNAC) thioesters were used.[20] We synthesized glycine SNAC (**5**) as a mimic of the proposed intermediate **I** (Figures S3-5). For biochemical assays, **5** was incubated with LbmA-Ox and reduced nicotinamide adenine dinucleotide phosphate (NADPH) as FAD-reducing agent.[21] To indirectly follow the reaction, NADPH consumption was monitored by spectroscopic measurements at 340 nm. As expected, NADPH levels decreased in contrast to a boiled enzyme control assay (Figure 2B). Ultra-high performance liquid chromatography/high-resolution mass spectrometry (UHPLC-HRMS) analysis indicated a new product **6** with a mass difference of +14 Da compared to **5**, in the assay containing active enzyme, consistent with the addition of one oxygen and loss of two hydrogens in oxime formation (Figures 2A and C). No such product was detected in the boiled enzyme control (Figure 2A, Figure S6). Further analysis by tandem mass spectrometry (MS/MS) revealed fragment ions consistent with the expected compound (Figure S7). To determine whether the installed oxygen originates from water or molecular oxygen, we repeated the assay in an $^{18}O_2$-enriched atmosphere. Compared to the control experiment, a +2 Da shift corresponding to incorporation of ^{18}O from molecular oxygen was observed (Figure 2D). In contrast, no ^{18}O incorporation was observed in an assay using an $H_2^{18}O$ background. The incorporation of molecular oxygen is therefore consistent with LbmA-Ox acting as a monooxygenase on **5**.[22]
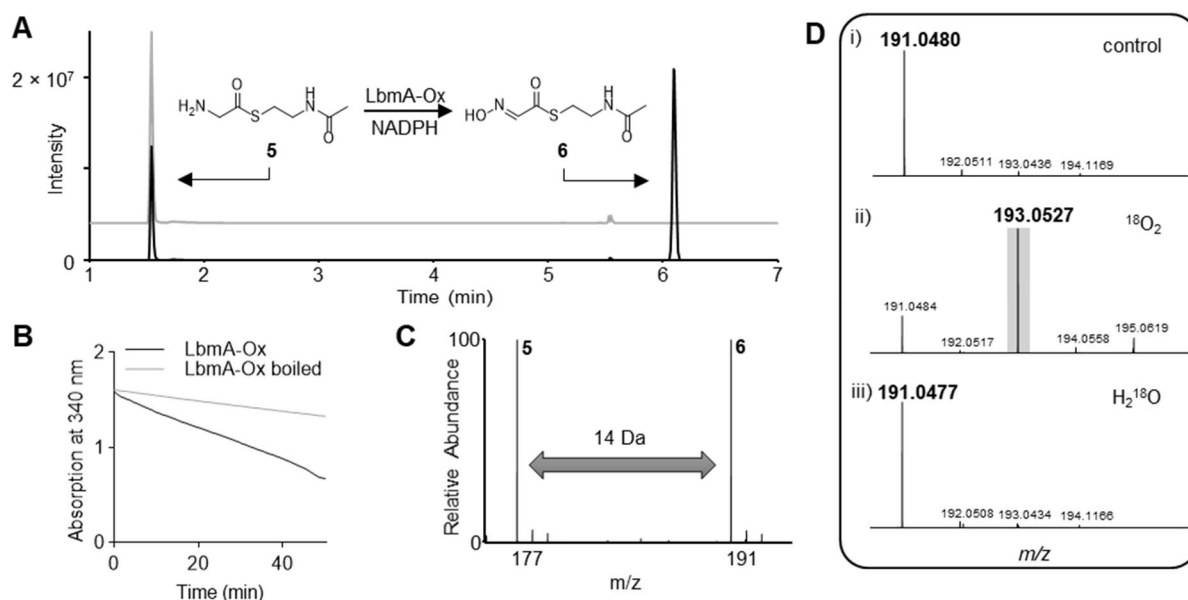
Figure 2. *In vitro* enzymatic assays using LbmA-Ox and thioester **5** as a test substrate. A) Extracted ion chromatogram (EIC) from UHPLC-HRMS data of assay mixtures for **5** and expected product **6** (calc. for [M+H]+ as 177.0692 and 191.0484 and measured as 177.0688 and 191.0481, respectively). In addition to the reactions with LbmA-Ox (black), negative control reactions containing boiled enzyme (grey) are shown. B) Absorption measurement at 340 nm to monitor NADPH conversion during the assay. C) Overlay of the MS spectra of **5** and **6** indicate a difference of 14 Da between substrate and product. D) Mass spectra of the putative product peak (**6**) and its natural isotope distribution for an assay under normal conditions (control), under $^{18}O_2$ enriched atmosphere ($^{18}O_2$), or in an $H_2^{18}O$ background ($H_2^{18}O$). The +2 Da shift corresponding to incorporation of ^{18}O is marked with a grey box.

We next aimed to explore the substrate scope of LbmA-Ox. Four further amino acid SNACs (**7**, **9-11**) were synthesized and analyzed in the same manner as **1** (Figure 3, Figures S8-25). For Ala-SNAC (**7**), we observed a product with *m/z* features and MS/MS fragments matching the corresponding monooxygenated compound (**8**, Figures S26-27). Only traces of the putative products were observed for test substrates with bulkier side chains (**9-11**, Figures 3 and S28) and no products were detected in assays using free amino acids as substrate (Figure S29). The lack of product formation for the latter indicates

the need for a carrier protein-bound substrate that is sufficiently mimicked by the SNACs. Isolation of the glycine thioester-derived product **6** for nuclear magnetic resonance (NMR) characterization was hampered by degradation under purification conditions. We measured the stability of the aminoacyl SNACs in hydrolysis assays using Ellman's reagent (Figure S30),[23] which confirmed thioester hydrolysis within minutes, as described previously for similar amino acid SNACs.[20d]
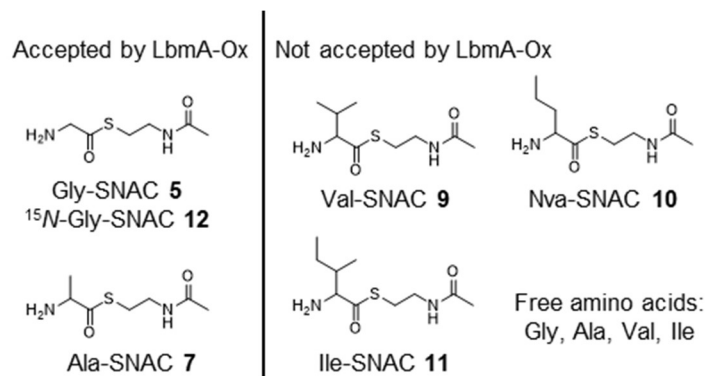
Figure 3. Test substrates for conversion with LbmA-Ox. Analytical data are shown in Figures 2 and S30-35.

In comparison to **7** and **9-11**, glycine SNAC **5** is particularly unstable. Correspondingly, a comparison of UHPLC-HRMS chromatograms before and after attempted isolations revealed the appearance of new peaks for the free SNAC thiol of hydrolyzed substrates or products (Figure S31). However, no putative glycine- or glyoxylate-like degradation products were detectable in positive or negative ionization mode. We therefore synthesized a ^{15}N-labelled glycine SNAC (**12**, Figures S32-36) as a substrate for assays that are subsequently analyzed by ^{15}N isotope-edited NMR. However, for unknown reasons, isotope-edited NMR experiments with parameters set for the detection of ^{1}H coupled to ^{15}N for either *E*- and *Z*-oxime moieties, imine, or amine moieties did not reveal detectable products. Likewise, in kinetic NMR experiments, no newly appearing signals corresponding to a product were detected (Figures S37-38). Instead, a new peak corresponding to an amide was observed. This might be explained by the rearrangement of the oxime to an amide via Beckmann rearrangement.[24]

To overcome these challenges, we investigated the role of LbmA-MT as a candidate for oxime *O*-methylation, which was hypothesized to convert ACP-tethered intermediate **II** to **III** (Figure 1A) with potentially increased stability. The corresponding MT gene region was cloned into pET28b-SUMO and expressed in *E. coli* as an N-terminally His$_6$- and SUMO-tagged protein. It was subsequently purified (Figure S39) and used in biochemical assays with MgCl$_2$, *S*-adenosyl methionine (SAM), and either **6** or **8** as test substrates, the latter obtained from assays with LbmA-Ox. Additionally, we conducted one-pot reactions with LbmA-Ox, LbmA-MT, coenzymes, and either **5** or **7**. However, no products corresponding to methylated **6** or **8** were detected by UHPLC-MS. Considering that LbmA-MT is module-embedded, we investigated its role *in vivo* within an intact PKS environment wherein native protein-protein interactions for LbmA-MT would be retained. We thus cloned the gene region encoding the first three modules of LbmA comprising the entire putative methyloxime-generating system, a ca. 3000 amino acid protein terminating with LbmA-KS3, into pET28b-SUMO and introduced the plasmid into *E. coli* BAP1. This strain harbors the *sfp* gene from *B. subtilis* encoding a promiscuous 4'-phosphopantetheinyl transferase (PPTase) for production of *holo* carrier proteins.[25] For the supply of glycine and cofactors, we relied on their natural occurrence in *E. coli* (Figure 4A). After production of the trimodular assembly line, the cells were lysed and intermediates were released from the protein by addition of cysteamine (Figure 4B).[26] UHPLC-HRMS analysis indicated formation of **II** and **III** observed as **13** (calc. for [C$_6$H$_{12}$N$_3$O$_2$S$_2$+H]$^+$ as 224.0522 and measured as 224.0519) and **14** (calc. for [C$_7$H$_{14}$N$_3$O$_2$S$_2$+H]$^+$ as 238.0678 and measured as 238.0674) with predicted molecular formulae

consistent with oxygenation and methylation (Figure 4C). These compounds were absent in a control experiment using empty pET28b-SUMO (Figure S40). Two peaks were observed for both, **13** and **14,** and may correspond to the *E* and *Z* isomers of the (methylated) oxime group. We only observed intermediates corresponding to a conversion of glycine but not of alanine. To further test whether the product contains a methyl group derived from SAM, we repeated the *in vivo* assay in medium supplemented with ^{13}C-methyl methionine. Detection of *m/z* features corresponding to **14** with a 1 Da mass increase supports incorporation of SAM originating from ^{13}C-methyl methionine (Figure S41).
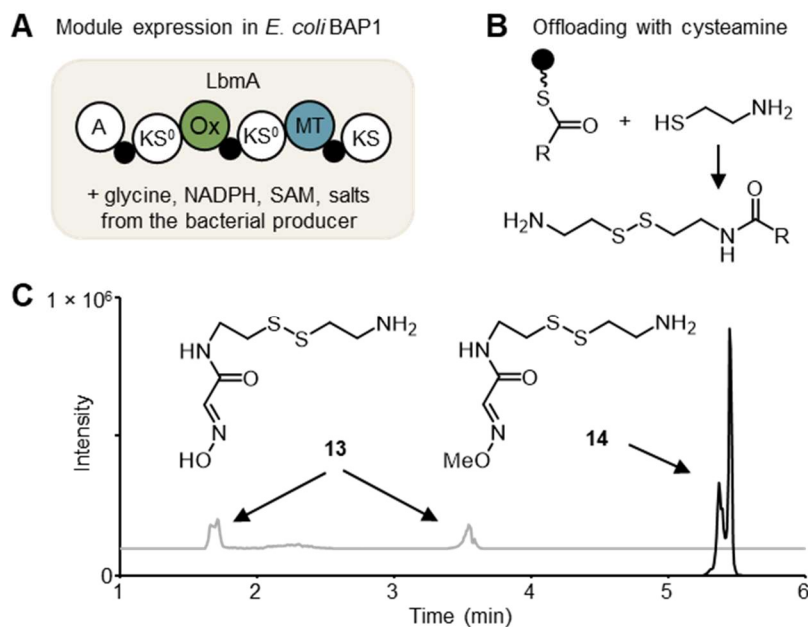
Figure 4. *In vivo* assay expressing the first three modules of LbmA in *E. coli* BAP1. Carrier proteins are activated to their *holo* form by the 4'-phosphopantetheinyl transferase Sfp. A) Expression strain harboring the first three modules on an expression vector. B) After expression and cell disruption, the cell lysate was treated with cysteamine. The offloaded intermediates were analyzed by UHPLC-HRMS. R: polyketide chain. C) Extracted ion chromatograms (EIC) from UHPLC-HRMS data. EICs for **13** (grey) and **14** (black) (calc. for [M+H]$^+$ as 224.0522 and 238.0678, and measured as 224.0519 and 238.0674, respectively). Data for control experiments are shown in Figures S40-41.

In organic extracts of lobatamide-producing *Gynuella sunshinyii*, no lobatamide congener biosynthesized from an alanine instead of a glycine incorporated at the loading module was detected. Accordingly, in the *in vivo* assay with adjacent LbmA-A and LbmA-KS^{0}1 present, only glycine-derived intermediates were observed. Since free LbmA-Ox accepts the non-native alanine thioester as a substrate in an *in vitro* setting, the adjacent A domain and perhaps the KS^0 domains might act as gatekeepers to prevent this reaction on the PKS. The signature of LbmA-A for glycine combined with the downstream LbmA-KS^{0}1 with a phylogenetic match to glycine-containing intermediates could ensure installation of exclusively glycine-derived oxime moieties in the elongated polyketide. Such a gatekeeping function has previously been attributed to amino-acid accepting KSs[27] and KS^{0}s,[28] but was not tested here. LbmA-KS^{0}2 may act similarly, keeping the unstable intermediate **II** protected while channeling it to LbmA-MT for methylation and stabilization.

Conclusion and Outlook

Reaction mechanisms for oxime-forming enzymes have been proposed for a number of natural product classes, most commonly involving *N*-hydroxylation by a cytochrome P450[29] or flavin-dependent monooxygenase.[30] For PKS, NRPS, or PKS/NRPS-hybrid pathways, post-assembly line, but not "on-line" oxime formation was reported for several natural products.[29a, 30d, 31] Comparison of the starting modules in the assembly lines of *trans*-AT PKS-derived V-ATPase inhibitors suggests a conserved trimodular system and mechanism for methyloxime formation, with the exception of a DH domain with unclear function in some PKSs. Taken together, our results support oxime installation as an integrated, modular function of *trans*-AT PKSs (Figure 5). Based on *in vitro* and *in vivo* assays, we propose the following order of events: After carrier protein loading by a glycine-specific A domain and transfer to the downstream PKS module, LbmA-Ox performs an *N*-hydroxylation of the carrier protein-bound amino acid via the incorporation of molecular oxygen. Subsequent oxime formation may be achieved either by direct dehydrogenation or by a second *N*-hydroxylation and presumably spontaneous dehydration. In the next module, LbmA-MT *O*-methylates the oxime. Due to the instability of oximes in aqueous solution, methylation may serve as protection from degradation.[29c] Our attempts to detect any putative dihydroxylated intermediate as well as derivatization of an aldehyde intermediate with phenylhydrazine were not successful. Previous studies on non-PKS-associated oxime-forming oxygenases with hydroxylated intermediates reported similar difficulties.[30a]
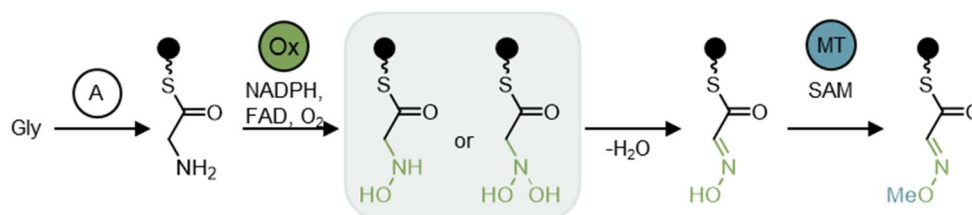
Figure 5. Putative mechanism of methylated oxime formation in *trans*-AT PKSs. Wavy lines indicate carrier protein-bound biosynthetic intermediates.

Trans-AT PKSs are fascinating enzymes with a biocatalytic scope that far exceeds the fatty acid synthase-type biochemistry of standard PKSs. To our knowledge, this is the first characterization of modular oxime formation for a PKS or NRPS system. Functional characterization of novel modules upstream of unassignable KSs are needed to expand the toolset for biosynthetic diversification and to improve the predictive power of tools like transATor in an iterative process. The confined, N-terminal module architecture and its functionality in *E. coli* make the oxime-forming modules an intriguing candidate for synthetic biology purposes.

Acknowledgements

References

[1]   M. R. Boyd, C. Farina, P. Belfiore, S. Gagliardi, J. W. Kim, Y. Hayakawa, J. A. Beutler, T. C. McKee, B. J. Bowman, E. J. Bowman, *J. Pharmacol. Exp.* **2001**, *297*, 114-120.

[2]   K. A. Dekker, R. J. Aiello, H. Hirai, T. Inagaki, T. Sakakibara, Y. Suzuki, J. F. Thompson, Y. Yamauchi, N. Kojima, *J. Antibiot.* **1998**, *51*, 14-20.

[3]   a) K. L. Erickson, J. A. Beutler, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **2001**, *66*, 1532-1532; b) K. L. Erickson, J. A. Beutler, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **1997**, *62*, 8188-8192.

[4]   D. L. Galinis, T. C. McKee, L. K. Pannell, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **1997**, *62*, 8968-8969.

[5]   a) B. Kunze, R. Jansen, F. Sasse, G. Höfle, H. Reichenbach, *J. Antibiot.* **1998**, *51*, 1075-1080; b) B. Kunze, H. Steinmetz, G. Höfle, M. Huss, H. Wieczorek, H. Reichenbach, *J. Antibiot.* **2006**, *59*, 664-668; c) Y. Hayakawa, T. Tomikawa, K. Shin-ya, N. Arao, K. Nagai, K. Suzuki, K. Furihata, *J. Antibiot. (Tokyo)* **2003**, *56*, 905-908; d) J. W. Kim, K. Shin-Ya, K. Furihata, Y. Hayakawa, H. Seto, *J. Org. Chem.* **1999**, *64*, 153-155; e) K. Suzumura, I. Takahashi, H. Matsumoto, K. Nagai, B. Setiawan, R. M. Rantiatmodjo, K. Suzuki, N. Nagano, *Tetrahedron Lett.* **1997**, *38*, 7573-7576.

[6]   M. Huss, H. Wieczorek, *J. Exp. Biol.* **2009**, *212*, 341-346.

[7]   a) R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem., Int. Ed.* **2020**, *59*, 7761-7765; b) R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem., Int. Ed.* **2020**, *59*, 11698.

[8]   S. P. Niehs, B. Dose, S. Richter, S. J. Pidot, H. M. Dahse, T. P. Stinear, C. Hertweck, *Angew. Chem., Int. Ed.* **2020**, *59*, 7766-7771.

[9]   a) M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468-3496; b) C. Hertweck, *Angew. Chem., Int. Ed.* **2009**, *48*, 4688-4716; cJ. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996-1047.

[10]   E. J. Helfrich, J. Piel, *Nat. Prod. Rep.* **2016**, *33*, 231-316.

[11]   a) D. E. Cane, C. T. Walsh, *Chem. Biol.* **1999**, *6*, R319-325; b) H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen, K. Sivonen, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 9259-9264.

[12]   a) E. J. N. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, M. H. Medema, *Nat. Commun.* **2021**, *12*, 1422; b) A. T. Keatinge-Clay, *Angew. Chem., Int. Ed.* **2017**, *56*, 4658-4660.

[13]   T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225-233.

[14]   E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, *Nat. Chem. Biol.* **2019**, *15*, 813-821.

[15]   M. Perez-Sayans, J. M. Somoza-Martin, F. Barros-Angueira, J. M. Rey, A. Garcia-Garcia, *Cancer Treat. Rev.* **2009**, *35*, 707-713.

[16]   H. Büttner, S. P. Niehs, K. Vandelannoote, Z. Cseresnyés, B. Dose, I. Richter, R. Gerst, M. T. Figge, T. P. Stinear, S. J. Pidot, C. Hertweck, *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2110669118.

[17]   R. D. Sussmuth, A. Mainz, *Angew. Chem., Int. Ed.* **2017**, *56*, 3770-3821.

[18]   M. Rottig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362-367.

[19] K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2021**, *49*, W29-W35.

[20] a) J. Franke, C. Hertweck, *Cell Chem. Biol* **2016**, *23*, 1179-1192; b) S. Yue, J. S. Duncan, Y. Yamamoto, C. R. Hutchinson, *J. Am. Chem. Soc.* **1987**, *109*, 1253-1255; c) D. E. Cane, C. C. Yang, *J. Am. Chem. Soc.* **1987**, *109*, 1255-1257; d) D. E. Ehmann, J. W. Trauger, T. Stachelhaus, C. T. Walsh, *Chem. Biol.* **2000**, *7*, 765-772.

[21] a) D. Holtmann, F. Hollmann, *ChemBioChem* **2016**, *17*, 1391-1398; b) S. Eswaramoorthy, J. B. Bonanno, S. K. Burley, S. Swaminathan, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9832-9837; c) E. Romero, M. Fedkenheuer, S. W. Chocklett, J. Qi, M. Oppenheimer, P. Sobrado, *Biochim. Biophys. Acta - Proteins Proteom.* **2012**, *1824*, 850-857.

[22] V. Massey, *J. Biol. Chem.* **1994**, *269*, 22459-22462.

[23] P. W. Riddles, R. L. Blakeley, B. Zerner, *Methods Enzymol.* **1983**, *91*, 49-60.

[24] M. Boero, T. Ikeshoji, C. C. Liew, K. Terakura, M. Parrinello, *J. Am. Chem. Soc.* **2004**, *126*, 6280-6286.

[25] a) L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585-1595; b) B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790-1792.

[26] K. Belecki, C. A. Townsend, *J. Am. Chem. Soc.* **2013**, *135*, 14339-14348.

[27] C. Kohlhaas, M. Jenner, A. Kampa, G. S. Briggs, J. P. Afonso, J. Piel, N. J. Oldham, *Chem. Sci.* **2013**, *4*, 3212-3217.

[28] H. Y. He, M. C. Tang, F. Zhang, G. L. Tang, *J. Am. Chem. Soc.* **2014**, *136*, 4488-4491.

[29] a) W. L. Kelly, C. A. Townsend, *J. Am. Chem. Soc.* **2002**, *124*, 8186-8187; b) O. Sibbesen, B. Koch, B. A. Halkier, B. L. Moller, *J. Biol. Chem.* **1995**, *270*, 3506-3511; c) B. M. Koch, O. Sibbesen, B. A. Halkier, I. Svendsen, B. L. Moller, *Arch. Biochem. Biophys.* **1995**, *323*, 177-186.

[30] a) Y. Zhu, Q. Zhang, S. Li, Q. Lin, P. Fu, G. Zhang, H. Zhang, R. Shi, W. Zhu, C. Zhang, *J. Am. Chem. Soc.* **2013**, *135*, 18750-18753; b) M. N. Goettge, J. P. Cioni, K. S. Ju, K. Pallitsch, W. W. Metcalf, *J. Biol. Chem.* **2018**, *293*, 6859-6868; c) N. Liu, L. Song, M. Liu, F. Shang, Z. Anderson, D. J. Fox, G. L. Challis, Y. Huang, *Chem. Sci.* **2016**, *7*, 482-488; d) A. Becerril, I. Perez-Victoria, S. Ye, A. F. Brana, J. Martin, F. Reyes, J. A. Salas, C. Mendez, *ACS Chem. Biol.* **2020**, *15*, 1541-1553.

[31] a) I. Garcia, N. M. Vior, A. F. Brana, J. Gonzalez-Sabin, J. Rohr, F. Moris, C. Mendez, J. A. Salas, *Chem. Biol.* **2012**, *19*, 399-413; b) I. Garcia, N. M. Vior, J. Gonzalez-Sabin, A. F. Brana, J. Rohr, F. Moris, C. Mendez, J. A. Salas, *Chem. Biol.* **2013**, *20*, 1022-1032.

Experimental Procedures

General

LC-ESI mass and MS/MS spectrometry were performed on a Thermo Scientific Q Exactive mass spectrometer coupled to a Dionex Ultimate 3000 UPLC system. NMR spectra were recorded on a Bruker BBO 400MHz S1 and a Bruker BBO 500MHz S2 without a cold probe were used for the analysis of the synthetic intermediates. Chemical shifts are given in parts per million (ppm) and were referenced to the solvent peaks at δh 2.50 and δc 39.51 for DMSO-d_6.[1] Multiplicities are given as follows: s: singlet, d: doublet, t: triplet, q: quartet, quint.: quintet, m: multiplet. The obtained data were processed and analysed with Bruker Topspin 3.5 and 4.0 software, as well as Mestrelab Research S.L. MestReNova 14.2.0 software. Chemicals were purchased by Sigma-Aldrich (subsidiary of Merck KGaA) if not stated otherwise. Chemicals and solvents were purchased from commercial suppliers and were used without further purification. For silica gel chromatography, distilled technical grade solvents and silica gel SilicaFlash® P60 (Silicycle) were used. Thin layer chromatography (TLC) was performed using aluminum sheets "TLC Silica gel 60 F254" from Merck Millipore® and analysed with UV-light, ninhydrine or by permanganate staining. Enzymes were purchased by New England Biolabs or Thermo Fisher. Protino Ni-NTA agarose and DNA purification kits were purchased form Macherey-Nagel.

Phylogenetic analysis of lobatamide KS sequences

The amino acid sequences of the KS from the *lbm* cluster were aligned to 1121 KS sequences from an in-house collection of annotated *trans*-AT PKS clusters with assigned biosynthetic models using the MUSCLE algorithm with default settings.[2] KS3 and KS5 from the erythromycin *cis*-AT PKS were used as outgroup. A phylogenetic tree was build using FastTree (version 2.1.10 +SSE3 +OpenMP, 16 threads, default settings)[3] and bootstrap analysis was performed with 1000 pseudo-replicate sequences. The phylogenetic tree was visualized using FigTree (version 1.4.3).[4]

Construction of expression plasmids and strains

The sequences encoding LbmA-Ox and LbmA-MT were amplified from *G. sunshinyii* YC6258 liquid culture using the primer pairs LbmA-Ox-F/R and LbmA-MT-F/R (Table S1). The fragments were gel purified, digested with *Spe*I and *Not*I or *Nde*I and *Sac*I and cloned into derivatives of pET28b, namely pET28b-SUMO_Ser or pET28HB-TS, yielding plasmids pET28b-SUMO-*lbmA*_Ox and pET28b-SUMO-*lbmA*_MT. The sequence encoding the LbmA bimodule was amplified from *G. sunshinyii* YC6258 as two parts using the primer pairs LbmA-M3.2-F/R and LbmA-M3.3-F/R. The pET28b-SUMO_Ser backbone was amplified using the primer pair LbmA-M3.1-F/R with overhangs to the LbmA bimodule. The fragments were gel purified and the LbmA-bimodule was assembled to pET28b-SUMO by Gibson assembly[5] to yield pET28b-SUMO-*lbmA*_begLob. *E. coli* DH5α was transformed with the final constructs, the plasmid re-isolated and introduced into *E. coli* DE3 Tuner (NEB) for expression of N-terminally His$_6$- and SUMO-tagged LbmA-Ox, N-terminally and N-terminally His$_6$- and SUMO-tagged LbmA-MT or into *E. coli* BAP1 for expression of the LbmA-bimodule. As negative control, *E. coli* BAP1 was also transformed with empty pET28b-SUMO_Ser.

Heterologous gene expression and protein purification

LB medium containing 50 μg/mL kanamycin was inoculated with the respective *E. coli* expression strain and incubated overnight at 37 °C. TB medium containing 0.4% glycerol and 50 μg/mL kanamycin was inoculated 1:100 and cultured at 37 °C until an OD$_{600}$ of 0.6-1.0. After cooling at 4 °C for 15 min, 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce gene expression. The culture was grown at 16 °C for 16-20 h and then harvested by centrifugation. Cell pellets were either frozen and stored at -80 °C or directly processed at 4 °C by resuspension in lysis buffer (50 mM phosphate buffer,

300 mM NaCl, 10% [v/v] glycerol). Sonication with a Sonicator Q700 (QSonica, Newton, USA) was used to disrupt the cells and the lysate was centrifuged to remove cell debris. The supernatant was incubated with Ni-NTA agarose (Macherey-Nagel, Oensingen, Switzerland) for 30 min at 4 °C and subsequently transferred onto a fretted column. The resin was washed once with 3 mL lysis buffer, once with 3 mL washing buffer (lysis buffer with 40 mM imidazole) and then eluted with elution buffer (lysis buffer with 250 mM imidazole). The proteins were desalted into lysis buffer using a PD MiniTrap™ (GE Healthcare) column and subsequently concentrated to 5-10 mg/mL using an Amicon® Ultra centrifugal filter (Merck Millipore). The presence of protein in the elution fractions was assessed by SDS-PAGE.

Synthesis of test substrates

S-(2-Acetamidoethyl) 2-aminoethanethioate; Gly-SNAC (**5**) and *S*-(2-Acetamidoethyl) 2-(amino-^{15}N)ethanethioate; ^{15}N-Gly-SNAC (**12**)

Coupling of *N*-(*tert*-butoxycarbonyl)glycine and *N*-acetylcysteamine was performed according to a previously published procedure to obtain *N*-(*tert*-butoxycarbonyl)glycine-SNAC.[6] At 0 °C, *N*-(*tert*-butoxycarbonyl)glycine (206.7 mg, 1.0 eq., 1.18 mmol) was dissolved in 5 mL dry DCM, followed by the addition of a catalytic amount of 4-dimethylaminopyridine. Then, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (226.2 mg, 1.3 eq., 1.45 mmol) was added and the mixture stirred for a few minutes. Finally, *N*-acetylcysteamine (132 μL, 1.1 eq., 1.24 mmol) was added and the reaction stirred overnight at room temperature. The reaction was quenched by addition of sat. aq. NH$_4$Cl (15 mL) and extracted twice with DCM. The combined organic phases were washed with water and dried over MgSO$_4$ and the organic phase was evaporated. For deprotection, 115 mg *N*-(*tert*-butoxycarbonyl)glycine-SNAC was dissolved in 4 mL 4 N dioxane-HCl on ice and stirred for 60 min. The solution was dried with compressed air and purified by flash chromatography using silica gel (5:1:1:1 butanol:water:acetic acid:acetone to 4:1:1:1 butanol:water:acetic acid:acetone).

5 (70.7 mg, 0.40 mmol) was isolated as a white solid with a yield of 34%. HRMS (*m/z*): [M+H]$^+$ calcd for C$_6$H$_{12}$N$_2$O$_2$S, 177.0692; found, 177.0693. ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.42 (s, br, 3H, NH$_2$), 8.15 (t, *J* = 5.60 Hz, 1H. NH), 4.06 (s, 2H, C$_α$H$_2$), 3.22 (q, *J* = 6.50 Hz, 2H, CH$_2$NH), 3.03 (t, *J* = 6.80 Hz, 2H, SCH$_2$), 1.80 (s, 3H, COCH$_3$). ^{13}C NMR (500 MHz, DMSO): δ 193.6 (C$_α$C=O), 169.4 (NHC=O), 47.1 (C$_α$), 37.9 (CH$_2$NH), 28.1 (SCH$_2$), 22.6 (NHCH$_3$).

12 was synthesized as described for **5**, starting with *N*-(*tert*-butoxycarbonyl)glycine-^{15}N. **12** (77.4 mg, 0.44 mmol) was isolated with slight impurities as a white solid with a yield of 37%. HRMS (*m/z*): [M+H]$^+$ calcd for C$_6$H$_{12}$NO$_2$S^{15}N, 178.0663; found, 178.0659. ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.04 (t, *J* = 5.40 Hz, 1H. NH), 7.68 (t, *J* = 6.20 Hz, 1H, NH$_2$), 7.45 (t, *J* = 6.20 Hz, 1H, NH$_2$), 3.81 (d, *J* = 6.20 Hz, 2H, C$_α$H$_2$), 3.14 (m, 2H, CH$_2$NH), 2.87 (t, *J* = 6.60 Hz, 2H, SCH$_2$), 1.78 (s, 3H, COCH$_3$). ^{13}C NMR (400 MHz, DMSO): δ 199.4 (C$_α$C=O), 169.3 (NHC=O), 50.1 (C$_α$), 38.7 (CH$_2$NH), 27.9 (SCH$_2$), 23.1 (NHCH$_3$).

S-(2-Acetamidoethyl) 2-aminopropanethioate; Ala-SNAC (**7**), *S*-(2-Acetamidoethyl) 2-amino-3-methylbutanethioate; Val-SNAC (**9**), *S*-(2-Acetamidoethyl) 2-aminopentanethioate; Nva-SNAC (**10**), and *S*-(2-Acetamidoethyl) 2-amino-3-methylpentanethioate; Ile-SNAC (**11**)

Synthesis of **7** and **9-11** was performed according to previously published procedures with slight modifications.[7] At room temperature, Boc-*N*-l-amino acid (1.06 mmol, 1.0 eq.) was dissolved in 4-5 mL dry DCM. After the addition of diisopropylethylamine (740 μL, 4.24 mmol, 4.0 eq.) and PyBOP (1.1 g, 2.12 mmol, 2.0 eq.), *N*-acetylcysteamine (0.12 mL, 1.16 mmol, 1.1 eq.) was added and the reaction was stirred overnight. The solvent was removed in vacuo and the crude mixture taken up in

25 mL of ethyl acetate. The organic layer was washed with 5% $KHSO_4$, 5% $NaHCO_3$ and brine before being dried over $MgSO_4$. The solvent was evaporated and the resulting oil directly subjected to deprotection. At 0 °C, the residue was dissolved in 1 mL DCM and 1 mL TFA was added drop-wise. The reaction was allowed to warm to room temperature and stirred overnight. Then, the TFA was removed *in vacuo* and the residue purified by silica column (5:1:1:1 butanol/water/acetic acid/acetone to 4:1:1:1 butanol/water/acetic acid/acetone) to give the resulting SNAC conjugates as yellow or colorless oils.

7 (141 mg, 0.64 mmol) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 59%. HRMS (*m/z*): $[M+H]^+$ calcd for $C_7H_{14}N_2O_2S$, 191.0849; found, 191.0845. ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.44 (s, br, 3H, NH_2), 8.10 (t, $J = 5.50$ Hz, 1H. NH), 4.31 (m, 1H, $C_αH$), 3.23 (m, 2H, CH_2NH), 3.03 (m, 2H, SCH_2), 1.79 (s, 3H, $COCH_3$), 1.44 (d, $J = 7.20$ Hz, 3H, $C_βH_3$). ^{13}C NMR (500 MHz, DMSO-d_6): δ 197.3 ($C_αC=O$), 169.4 (NHC=O), 54.5 ($C_α$), 37.8 (CH_2NH), 28.3 (SCH_2), 22.6 ($NHCH_3$), 17.0 ($C_β$).

9 (124 mg, 0.572) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 54%. HRMS (*m/z*): $[M+H]^+$ calcd for $C_9H_{18}N_2O_2S$, 219.1162; found, 219.1155. ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.44 (s, br, 2H, NH_2), 8.09 (t, $J = 5.40$ Hz, 1H. NH), 4.15 (m, 1H, $C_αH$), 3.23 (m, 2H, CH_2NH), 3.06 (m, 2H, SCH_2), 2.18 (m, 1H, $C_βH$), 1.79 (s, 3H, $COCH_3$), 0.97 (dd, $J = 6.95$, 14.20 Hz, 6H, $C_β(CH_3)_2$). ^{13}C NMR (400 MHz, DMSO-d_6): δ 196.2 ($C_αC=O$), 169.4 (NHC=O), 63.5 ($C_α$), 37.8 (CH_2NH). 30.1 ($C_β$), 28.4 (SCH_2), 18.1, 17.3 ($C_γ$).

10 (141 mg, 0.64 mmol) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 61%. HRMS (*m/z*): $[M + H]^+$ calcd for $C_9H_{18}N_2O_2S$, 219.1162; found, 219.1156. ^{1}H NMR (400 MHz, DMSO-d_6): δ 8.48 (s, br, 3H, NH_2), 8.09 (t, $J = 5.60$ Hz, 1H. NH), 4.25 (t, $J = 6.30$ Hz, 1H, $C_αH$), 3.22 (m, 2H, CH_2NH), 3.05 (m, 2H, SCH_2), 1.79 (s, 3H, $COCH_3$), 1.77 (m, 2H, $C_βH_2$), 1.36 (m, 2H, $C_γH_2$), 0.98 (t, $J = 7.30$ Hz, 6H, $C_δH_3$). ^{13}C NMR (400 MHz, DMSO-d_6): δ 196.8 ($C_αC=O$), 169.4 (NHC=O), 58.0 ($C_α$), 37.6 (CH_2NH). 33.0 ($C_β$), 27.9 (SCH_2), 17.2 ($C_γ$), 13.3 ($C_δ$).

11 (128 mg, 0.55 mmol) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 52%. HRMS (*m/z*): $[M + H]^+$ calcd for $C_{10}H_{20}N_2O_2S$, 233.1318; found, 233.1318. ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.48 (s, br, 2H, NH_2), 8.10 (s, br, 1H. NH), 4.19 (m, 1H, $C_αH$), 3.24 (m, 2H, CH_2NH), 3.07 (m, 2H, SCH_2), 1.91 (m, 1H, $C_βH$), 1.80 (s, 3H, $COCH_3$), 1.47 (m, 1H, $C_γH$), 1.27 (m, 1H, $C_γH$), 0.95 (d, $J = 6.91$ Hz, 3H, $C_βCH_3$), 0.90 (t, $J = 4.25$ Hz, 3H, $C_δH_3$). ^{13}C NMR (500 MHz, DMSO-d_6): δ 195.9 ($C_αC=O$), 169.4 (NHC=O), 62.7 ($C_α$), 37.8 (CH_2NH). 36.5 ($C_β$), 28.5 (SCH_2), 24.3 ($C_γ$), 14.3 ($C_βCH_3$), 11.5 ($C_δ$).

In vitro enzyme activity assay

The assay was set up in a volume of 100 μL, containing purified LbmA-Ox (2-5 μM), substrate (2-5 mM), 1 mM NADPH, 2.5% [v/v] glycerol, and 50 mM phosphate buffer pH 6. Negative controls included boiled LbmA-Ox. The reactions were incubated for 45 min at 25 °C. Absorption at 340 nm was measured on an Infinite M200 Pro (Plex) Tecan spectrometer with 2 scans per second. Assays with loaded LbmA-ACP were incubated with 0.2 M cysteamine hydrochloride for 30 min at 30 °C prior to extraction. Reactions were extracted with ethyl acetate, dried, and resuspended in methanol. The samples were centrifuged at maximum speed for 10 min and the supernatant was subjected to UHPLC-HRMS analysis. For assays in $H_2^{18}O$, assay buffer was lyophilized and stored at -20 °C. For the assay, the lyophilized pellet was resuspended in $H_2^{18}O$. For assays under $^{18}O_2$, the buffer was degassed and $^{18}O_2$ was added. The remaining assay components were added under oxygen-free conditions. Extraction and analysis were performed as described above. For one-pot LbmA-Ox/LbmA-MT assays, purified LbmA-MT (2-5 μM), 0.5 mM *S*-adenosyl methionine (SAM), 150 mM NaCl, and 5 mM $MgCl_2$ were added to the assay mixture. For cascaded LbmA-Ox/LbmA-MT assays, extracts from an LbmA-Ox assay were resuspended in water and used as substrate. The assay was set up in a volume of 100 μL,

containing purified LbmA-MT (2-5 μM), 2.5% [v/v] glycerol, 50 mM phosphate buffer pH 6, 0.5 mM SAM, 150 mM NaCl, and 5 mM MgCl₂. After incubation at 25 °C for 45 min, the reactions were extracted with ethyl acetate and subjected to UHPLC-HRMS analysis. To check for hydrolyzed products, the assay was kept at 25 °C for an additional 48 h.

In vivo assay

The LbmA bimodule was expressed as described above using TB medium containing 4% glycerol. For assays using labelled methionine, the medium was supplemented with 5 mM l-methionine-(*methyl*-^{13}C). The cell lysate after sonication was incubated with 0.2 M cysteamine hydrochloride for 30 min at RT and extracted with ethyl acetate. The organic phase was evaporated and the residue was resuspended in methanol. The samples were centrifuged at maximum speed for 15 min and the supernatant was subjected to LC-MS analysis.

UHPLC-HRMS analysis

Measurements were conducted on a QExactive Orbitrap MS (Thermo Scientific) coupled to an UltiMate 3000 UHPLC system (Dionex). A solvent gradient (A = H₂O + 0.1% formic acid and B = acetonitrile + 0.1% formic acid was used on a Kinetex 2.6 μm C18 100 Å, LC Column (150 × 4.6 mm; Phenomenex). The gradient was solvent B at 2% for 0-2 min, 2-60% for 2-7 min, 60-95% for 7-8 min, 95% for 8-10 min at a flow rate of 1.0 mL/min. MS measurements were conducted in positive ionization mode in a mass range of 50-400 *m/z*. The spray voltage was set to 3.7 kV and the capillary temperature to 320 °C. MS/MS data were acquired in a data-dependent fashion with the parent ion scan at a resolution of 70,000 and the MS/MS scan at a resolution of 17,500. The 10 most abundant peaks of each parent ion scan were subjected to CID fragmentation with a normalized collision energy of 15, 20 and 25 and the dynamic exclusion time was set to 10 sec. MS/MS scans were conducted with an AGC target of 3×10^6 or a maximum injection time of 150 ms. Collected data of the MS experiments was analyzed with the Thermo Xcalibur 2.2. software.

Kinetic and ^{15}N NMR analysis

NMR experiments were performed on a Bruker AVNEO 700 MHz spectrometer equipped with a CP-QCI H&F-N/C/P-D 05 Z probehead or on a Bruker AVNEO 500 MHz spectrometer equipped with a CP-TCI H-C/N-D 05 Z probe head using the topspin 4.0.7 software for data acquisition and processing. NMR data were recorded at 298 K in standard 5 mm NMR tubes. To follow the enzymatic reaction, a series of 1D ¹H experiments with a 3-9-19 watergate sequence[8] was applied. A total of 128 transients was averaged with number of datapoints for the FID set to 32768. Protons which are scalar coupled to ^{15}N are selectively detected with isotope edited experiments. Isotope edited experiments result in clean spectra with only the signals of interest from the ^{15}N labeled substrate or product observable. To observe protons which are scalar coupled to ^{15}N an isotope edited 1D experiment with watergate[9] for water suppression was performed. The experimental parameters were set to observe one of the following moieties: *E*- or *Z*-oxime, amide, imine, or amide protons coupled to [15]N. The INEPT delays for the isotope filters were set according to the coupling constant J_{H-N}. The carrier offset for ^{15}N was set to the reported values for ^{15}N in the corresponding functional group. A total of 128 scans was averaged for the kinetic experiments and 5120 transients were averaged at the end of the reactions with the size of the FID set to 2794 complex data points.

Chemical shifts and scalar coupling constants (values for ^{15}N are from Witanowski[10]):

functional group	$\delta(^{15}N)$ / ppm	$J_{H\text{-}N}$ / Hz	$\delta(^1H)$ / ppm
E-oxime	350	2.7	7 - 8
Z-oxime	350	13.9	7 - 8
amine	30	1	4 - 5
amide	120	90	7 - 8
imine	340	4	10
hydroxylamine	120	1	3

NMR samples containing the enzyme, NADPH and the substrate in assay buffer were prepared immediately before data acquisition. The first time point was acquired approximately 5 minutes after sample mixing and experiments were repeatedly acquired up to 10 hours after sample mixing. For each time point a 1D and an isotope edited 1D experiment were performed. The total experiment time for both experiments is about 20 minutes. At the end of the reaction, an isotope edited 1D experiments and 2D ^{15}N-^1H HSQC experiments employing watergate[9] water suppression were performed to identify possible reactions products.

Supplementary Tables and Figures

Table S1. Primers used in this work

Primer Name	Primer Sequence
LbmA-Ox-F	CGGACTAGTCCGGAAATCATCATTCTC
LbmA-Ox-R	TTGCGGCCGCTTTCAGATACGTTGC
LbmA-MT-F	CGGACTAGTCCGAGCGCGATGCCAGAGGCATTC
LbmA-MT-R	TTGCGGCCGCTTTGACAGCTCTGCCGCTGGGGATTC
LbmA-M3.1-F	ACCTGCTTGGCGAAATATCGCGGCCGCACTCGAGCACCACCACCACC ACC
LbmA-M3.1-R	TCGTTGAAAGAAAGTTTCATCTAGTGGATCCACCAATCTGTTCTCTGT GAGCCTCAATAATATCGTTATCCTCC
LbmA-M3.2-F	CAGATTGGTGGATCCACTAGATGAAACTTTCTTTCAACGAATTGTTAC ACTTAAAGAGCACGTTTC
LbmA-M3.2-R	TTTCATTTCCTCTTGTCTGTACGGGGTGGCCCAGTAGGCTTTCTGATC AA
LbmA-M3.3-F	AGCCTACTGGGCCACCCCGTACAGACAAGAGGAAATGAAACCAATG ACAGACC
LbmA-M3.3-R	GTGGTGCTCGAGTGCGGCCGCGATATTTCGCCAAGCAGGTACCGGGC AAG
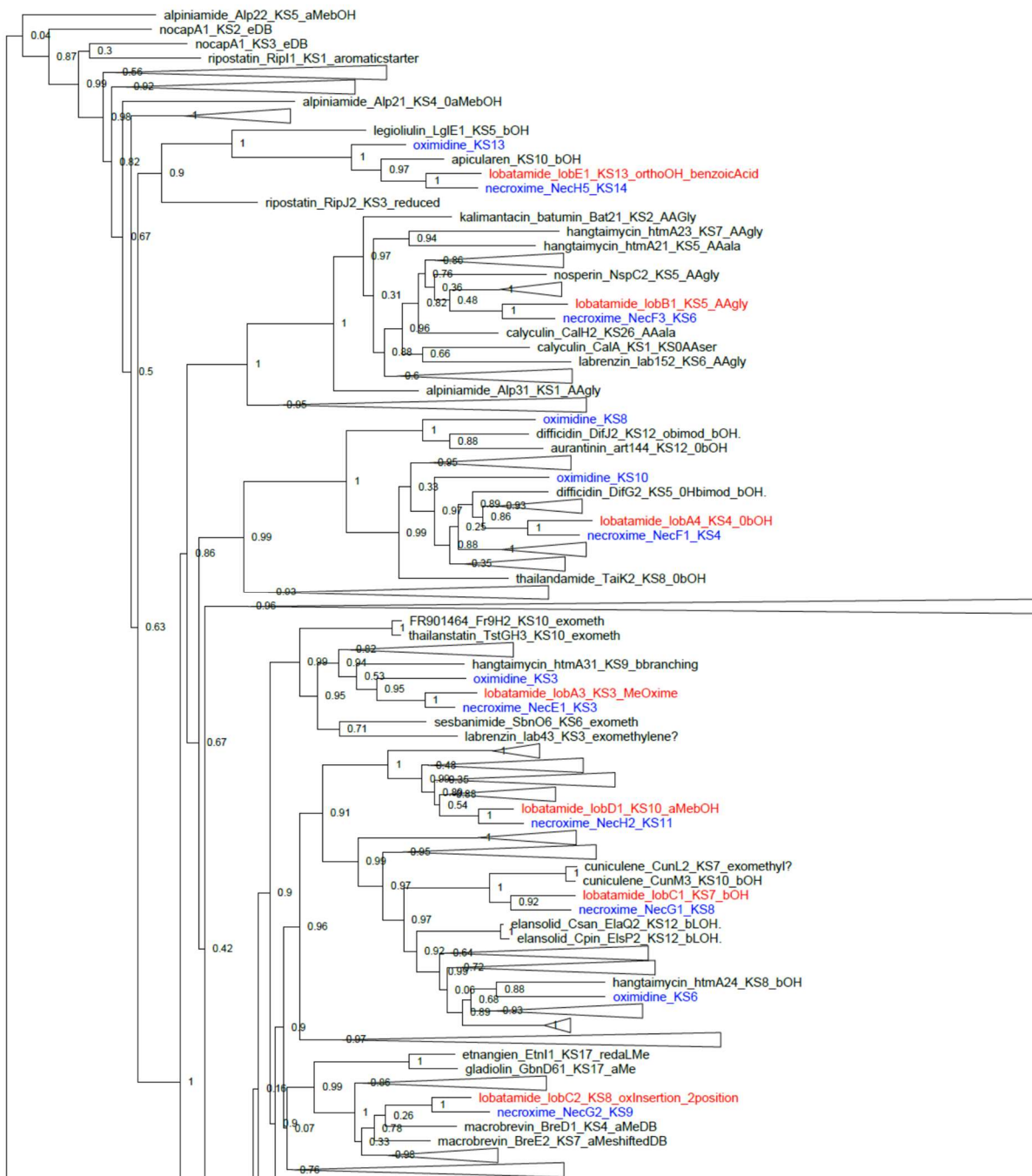
Figure S1. Maximum likelihood phylogram of 1134 KS domains from 88 *trans*-AT PKS clusters using two KSs from the *cis*-AT PKS for erythromycin as outgroup (bottom). KS sequences are named as compound_protein_number of KS_assigned biochemical transformation. Ketide clades for necroxime, oximidine, and lobatamide are expanded. Lobatamide KSs are marked in red, necroxime and oximidine KSs in blue. Nodes are labelled with their numerical bootstrap values.
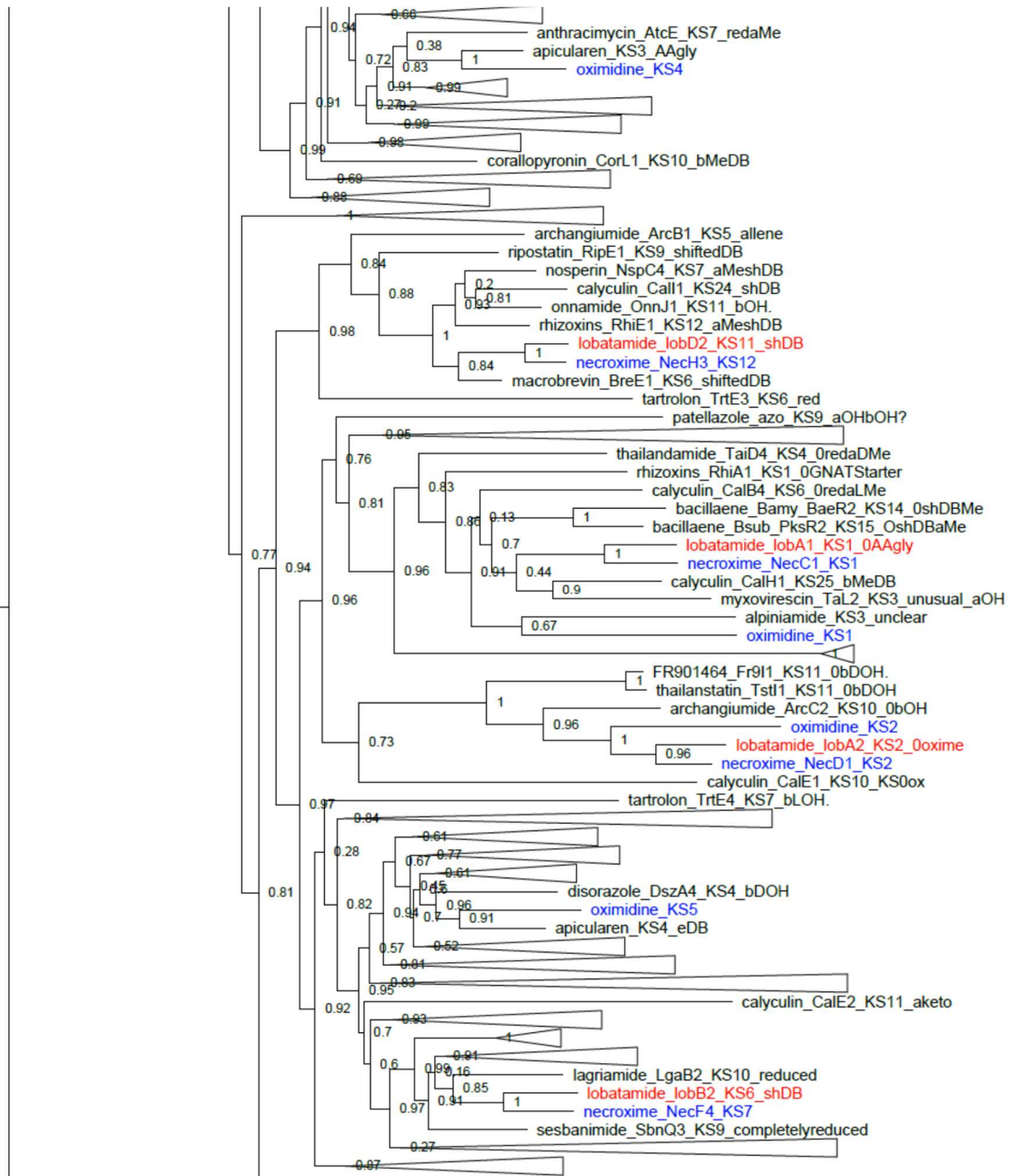
Figure S1 continued.
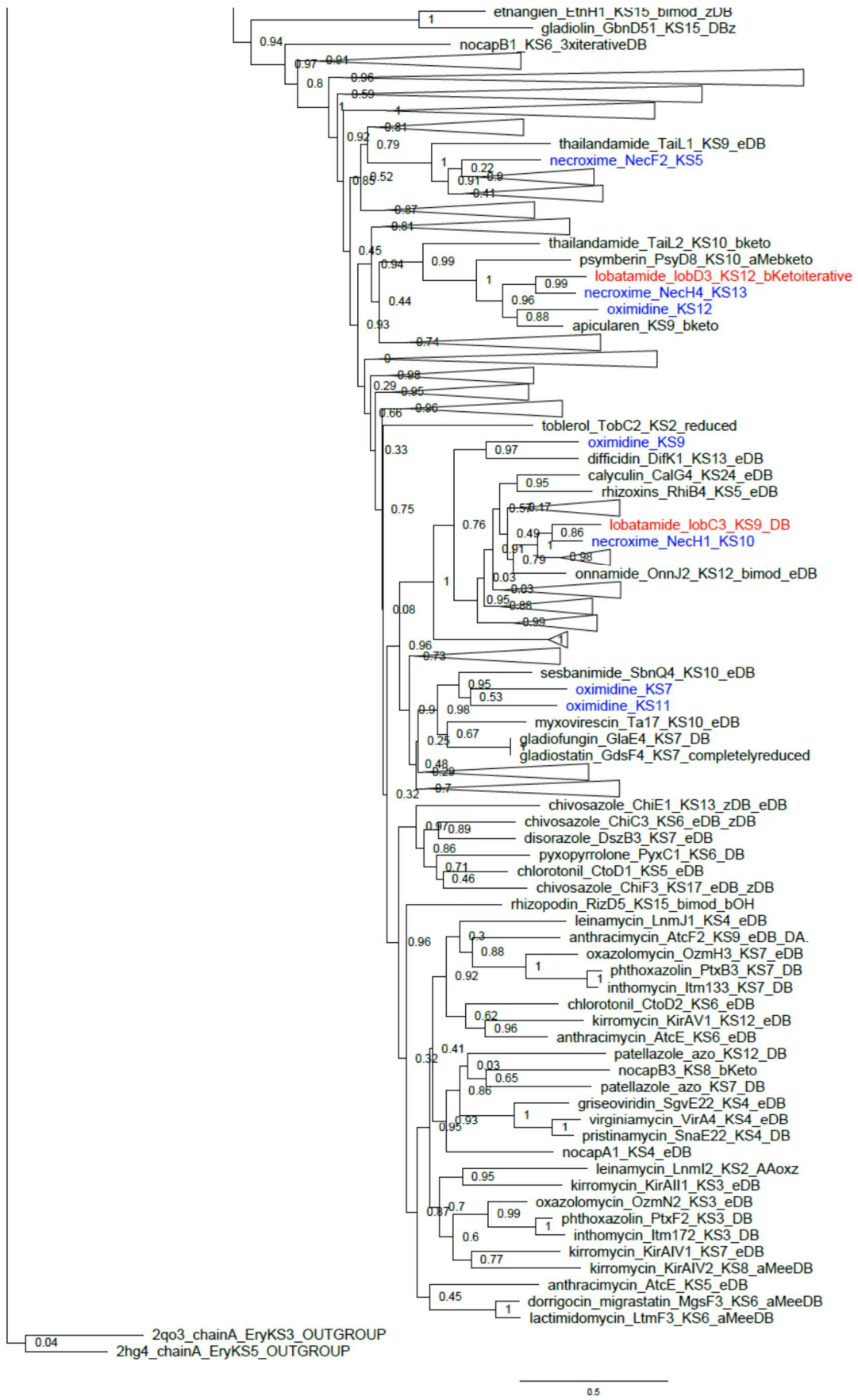
Figure S1 continued.
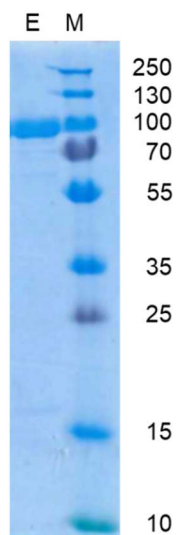
Figure S2. 12% SDS-PAGE gel of His$_6$-SUMO-tagged LbmA-Ox post Ni-NTA affinity purification. Expected molecular mass: 91.3 kDa. E: elution, M: protein ladder.
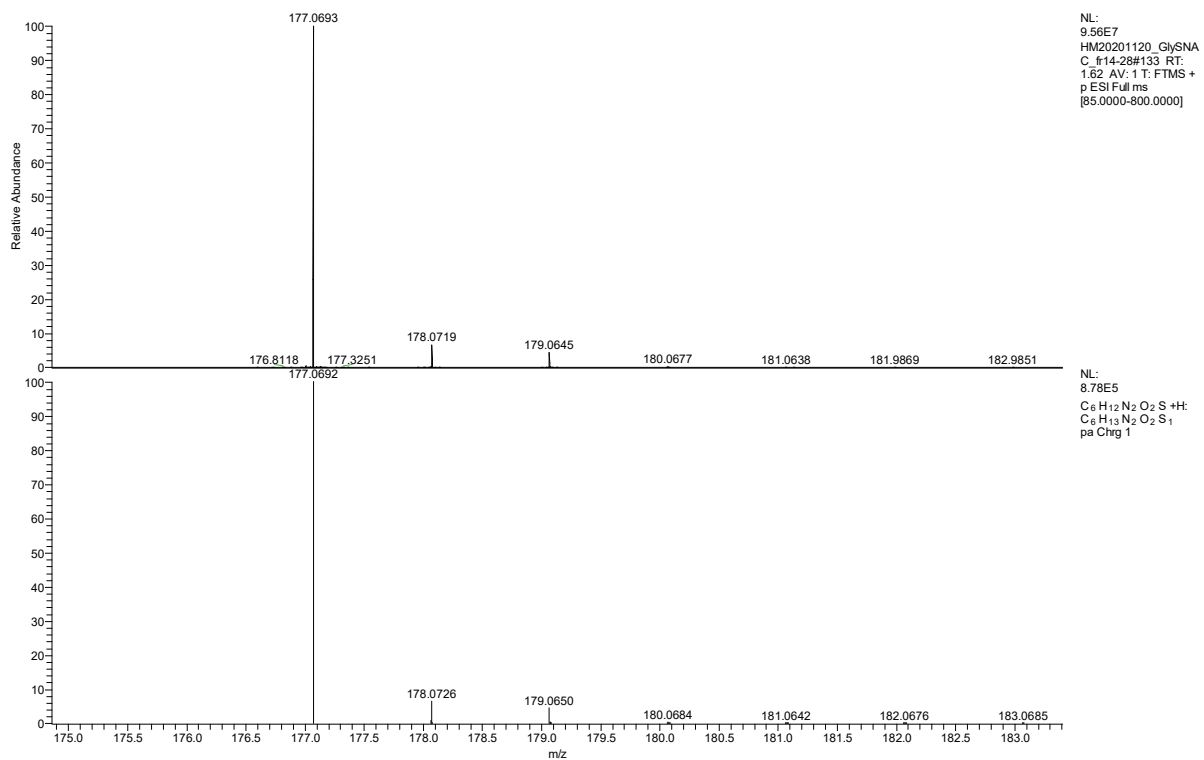
Figure S3. Top: mass spectrum for Gly-SNAC (**5**). Bottom: simulated mass spectrum.

NL:
9.56E7
HM20201120_GlySNAC_fr14-28#133 RT:
1.62 AV: 1 T: FTMS +
p ESI Full ms
[85.0000-800.0000]

NL:
8.78E5
$C_6 H_{12} N_2 O_2 S$ +H:
$C_6 H_{13} N_2 O_2 S_1$
pa Chrg 1
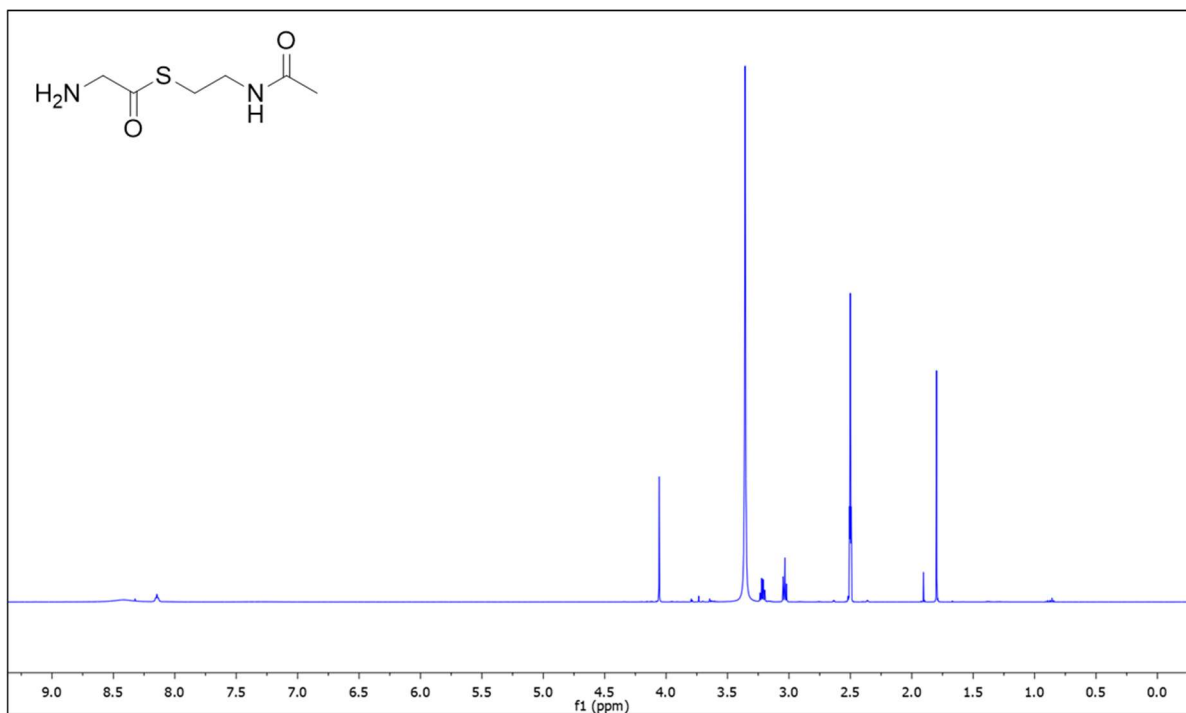
Figure S4 ^{1}H NMR spectrum of Gly-SNAC (**5**) in DMSO-d_6 at 500 MHz.
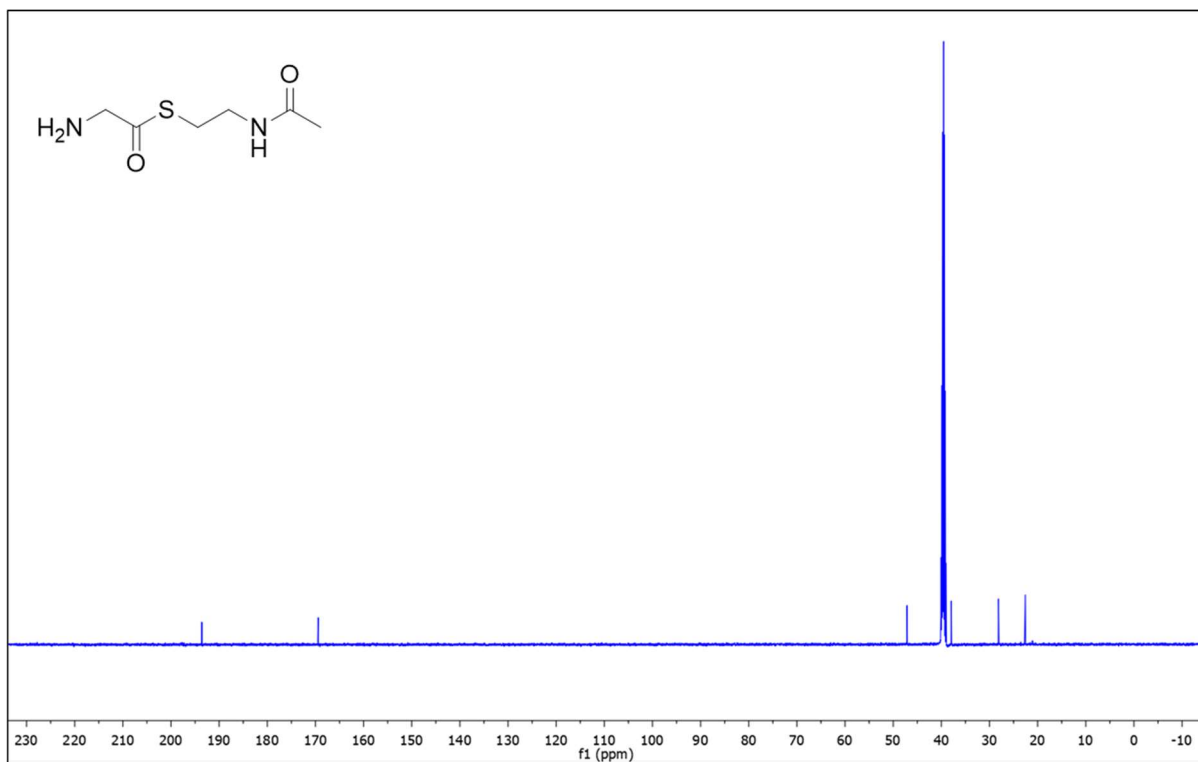
**Figure S5.** ^{13}C NMR spectrum of Gly-SNAC (**5**) in DMSO-d_6 at 500 MHz.
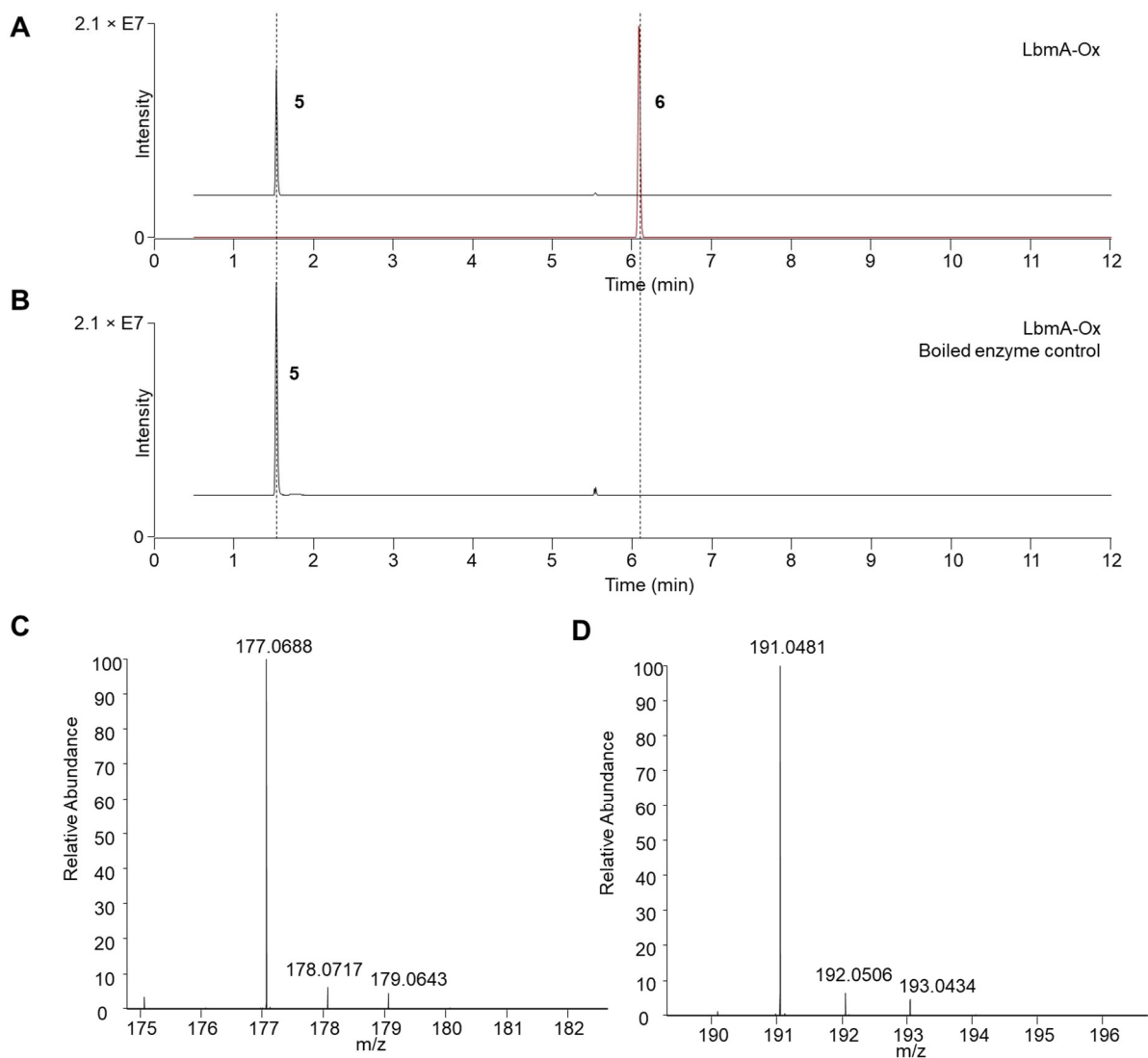
Figure S6. Extracted ion chromatogram (EIC) for *in vitro* LbmA-Ox assay mixtures using Gly-SNAC (**5**) as substrate. A) EIC for substrate **5** and product **6**. B) EIC for substrate **5** and product **6** for the control assay using boiled enzyme. C) Mass spectrum of **5** (calculated for [M+H]$^+$ 177.0692) at 1.54 min. D) Mass spectrum of **6** (calculated for [M+H]$^+$ 191.0484) at 6.09 min.
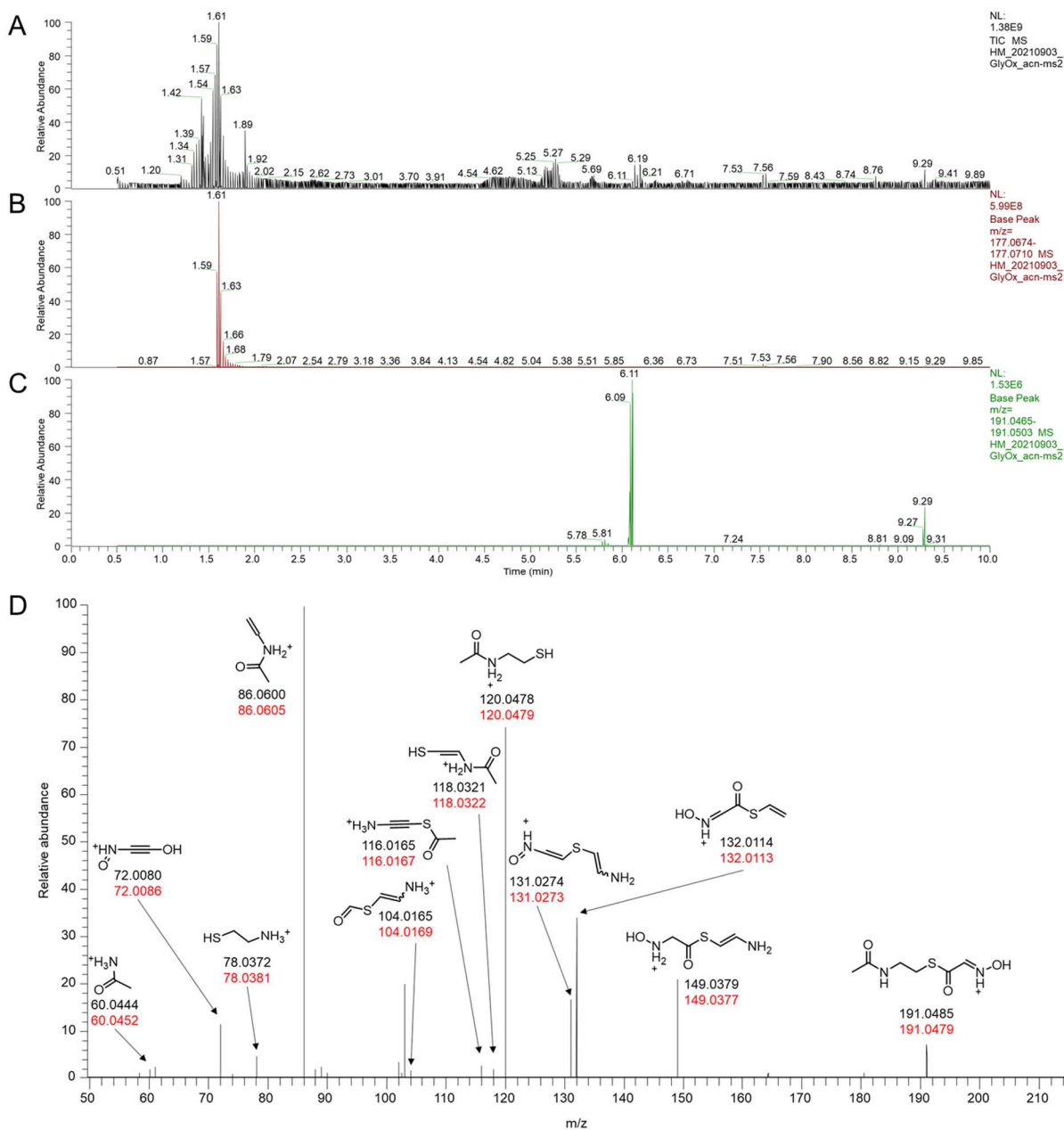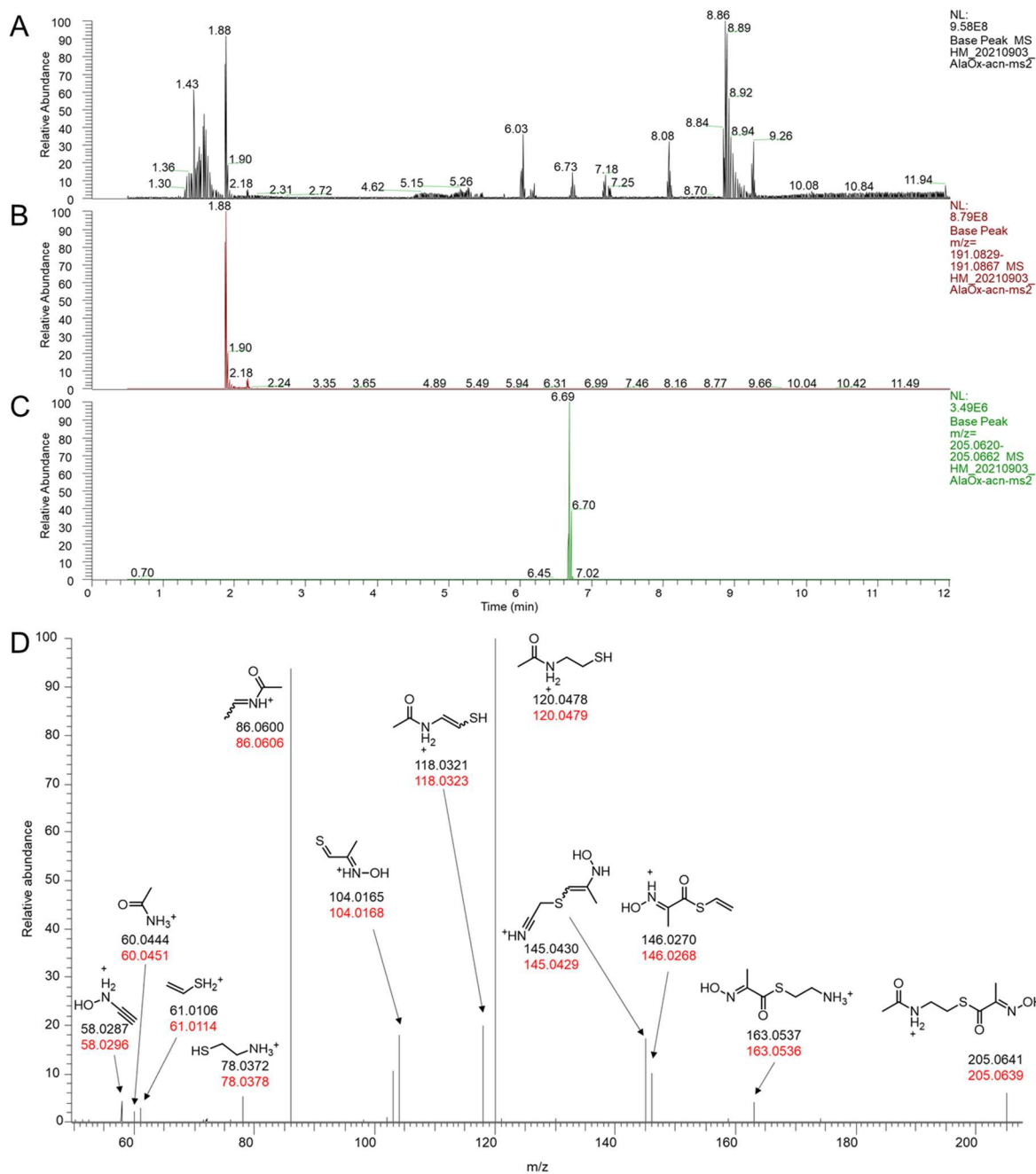
Figure S7. UHPLC-HRMS data of *in vitro* LbmA-Ox assay reaction with Gly-SNAC (**5**) as substrate. A) Total ion chromatogram. B) Extracted ion chromatogram for substrate **5**. C) Extracted ion chromatogram for product **6**. D) Filtered MS/MS spectrum of mass 191.0480 at 6.09 min. Black numbers indicate calculated *m/z* and red numbers indicate measured *m/z*. Possible fragment ions shown were deduced from predicted structures using the competitive fragmentation modeling for metabolite identification (CFM-ID 3.0) tool.[11]

Figure S8. Top: mass spectrum for Ala-SNAC (**7**). Bottom: simulated mass spectrum.
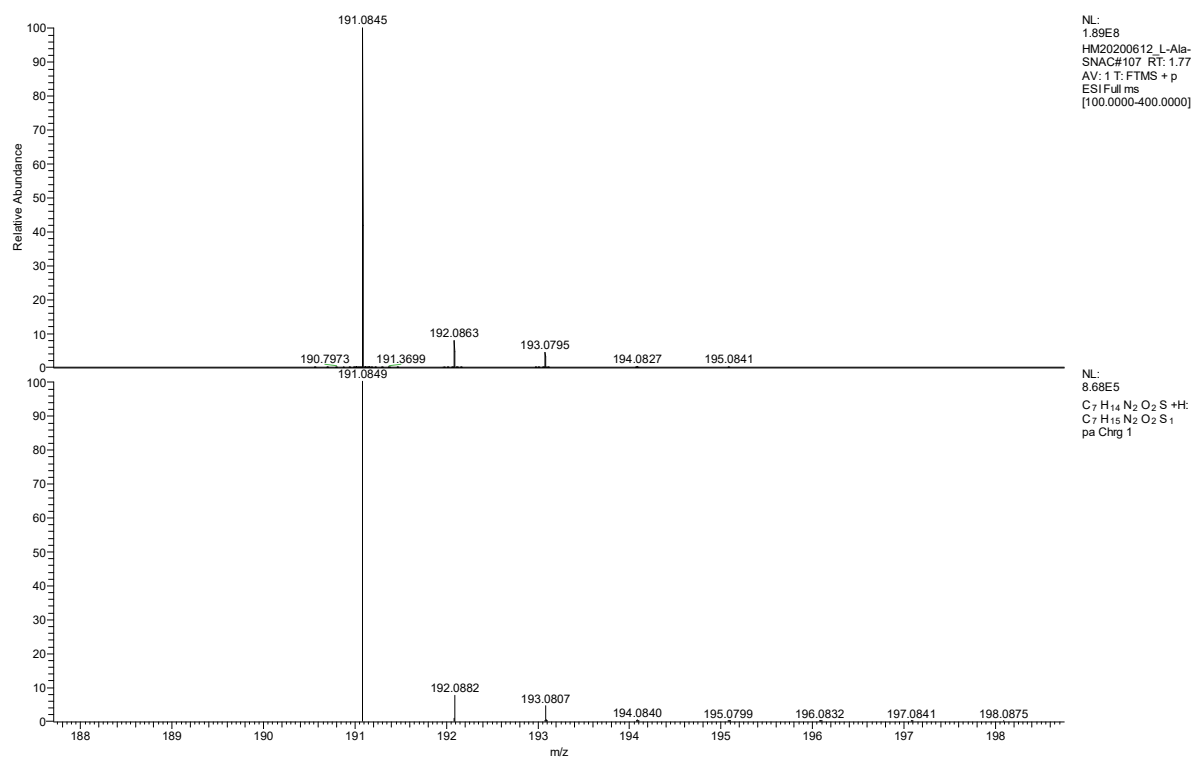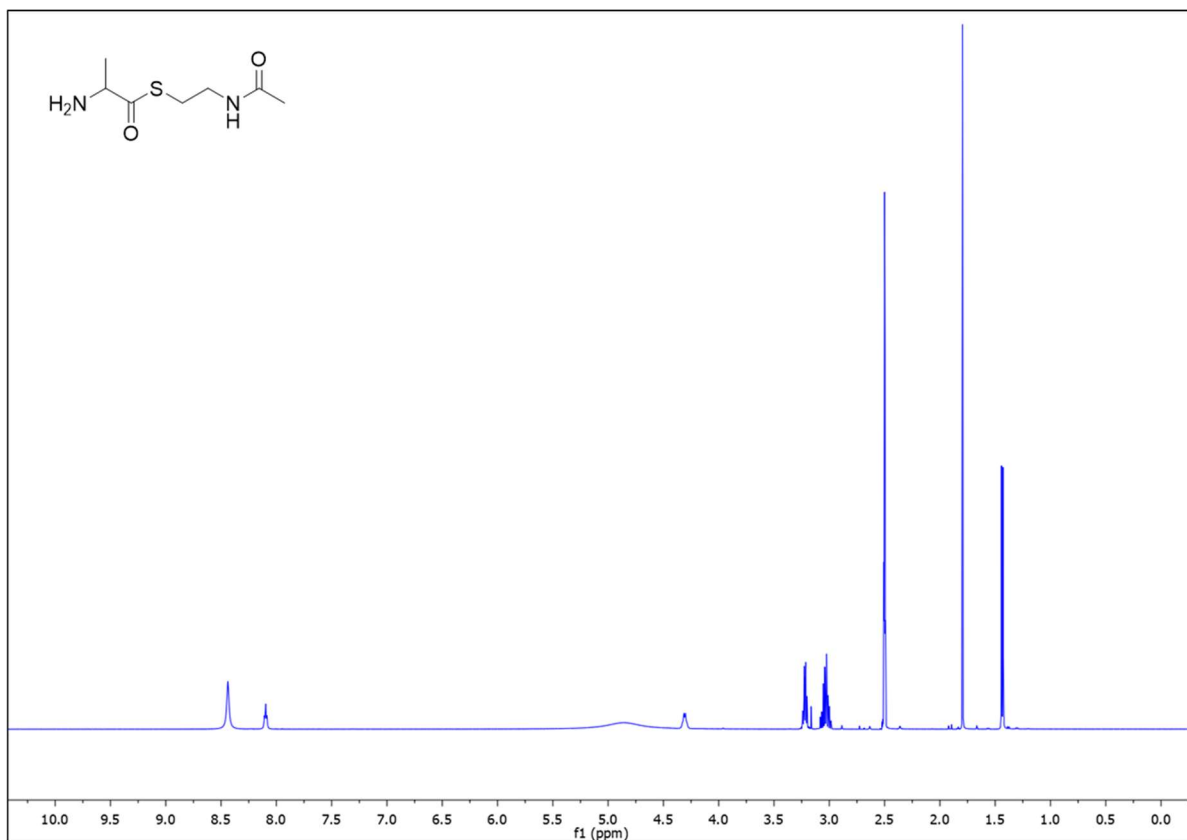
Figure S9. ^{1}H NMR spectrum of Ala-SNAC (**7**) in DMSO-d_6 at 500 MHz.
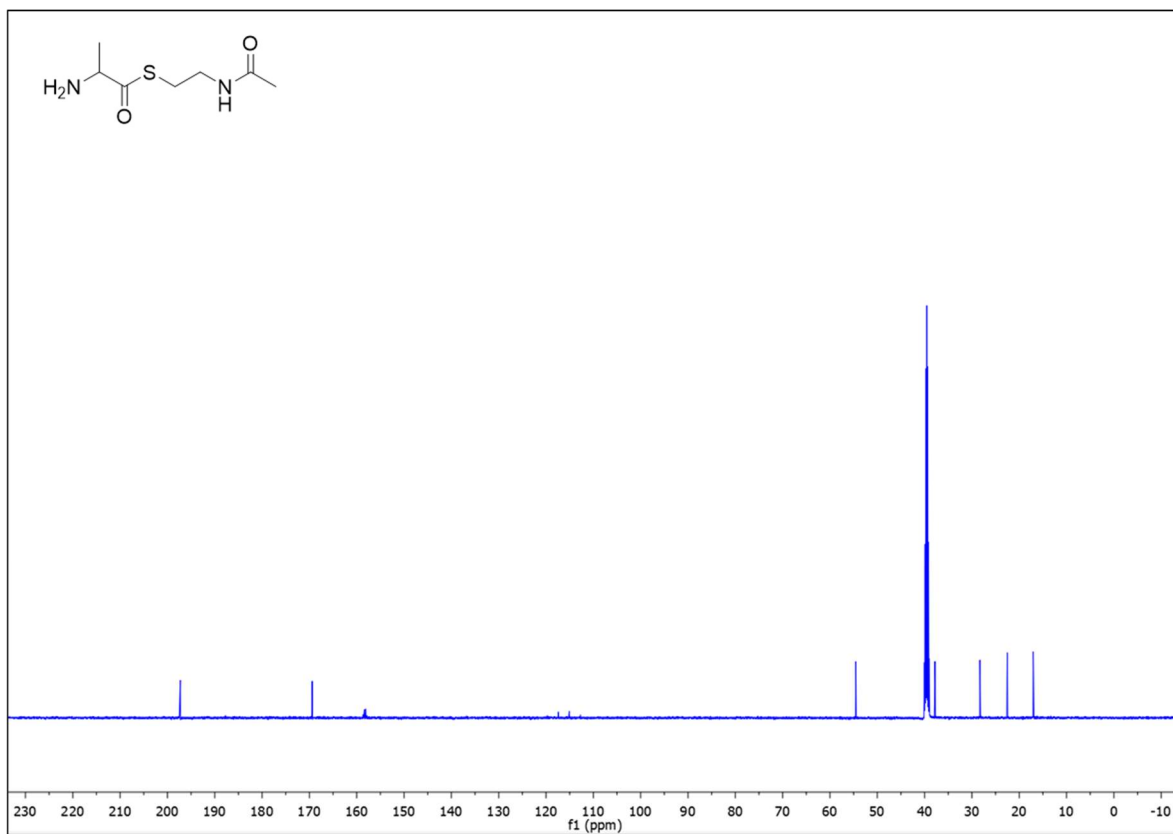
Figure S10. ^{13}C NMR spectrum of Ala-SNAC (**7**) in DMSO-d_6 at 500 MHz.
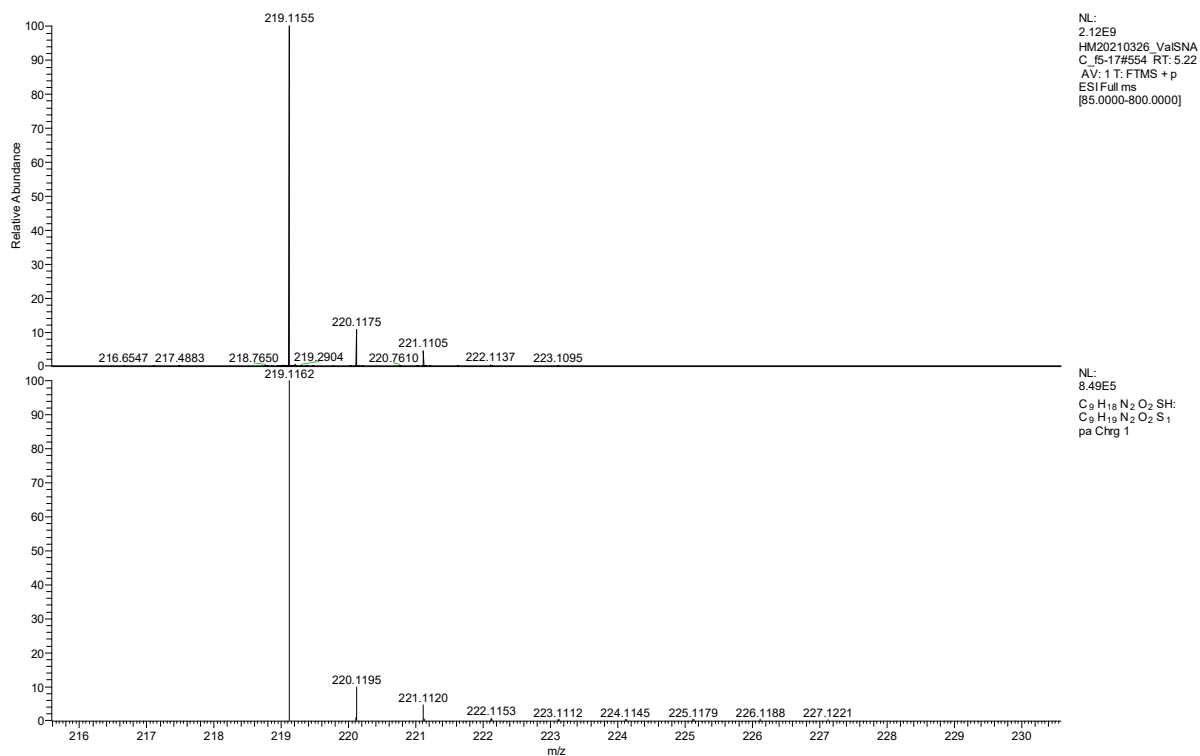
Figure S11. Top: mass spectrum for Val-SNAC (**9**). Bottom: simulated mass spectrum.
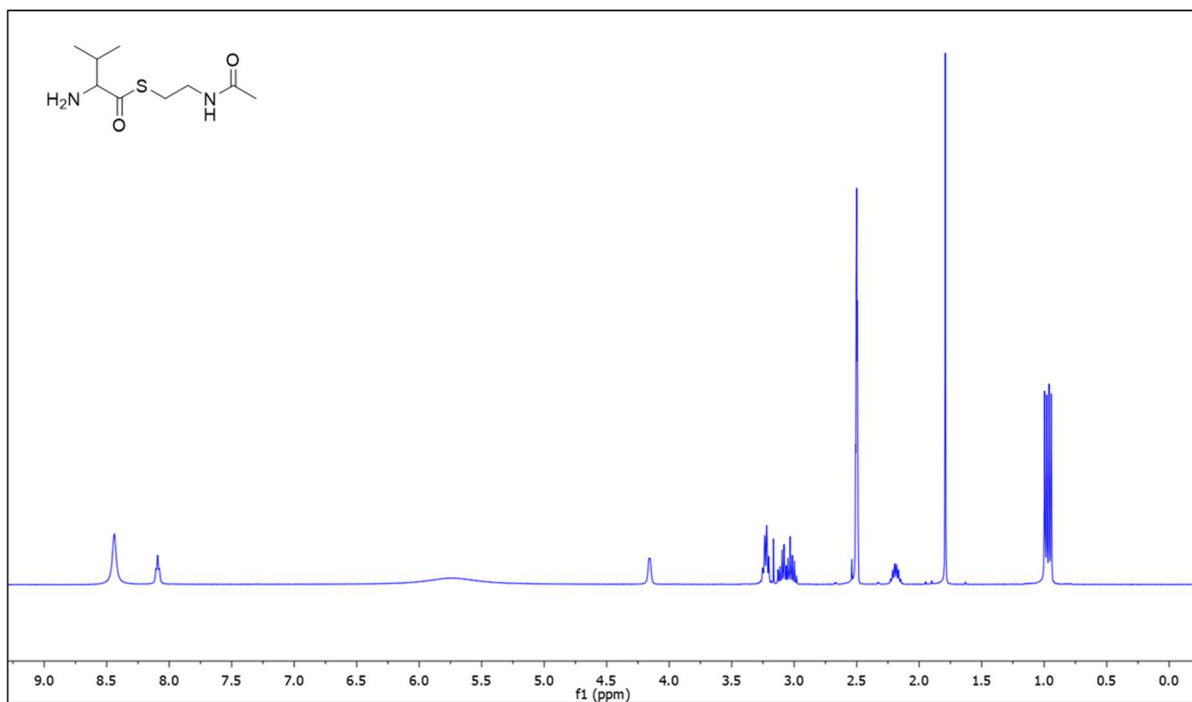
Figure S12. ^{1}H NMR spectrum of Val-SNAC (**9**) in DMSO-d_6 at 400 MHz.
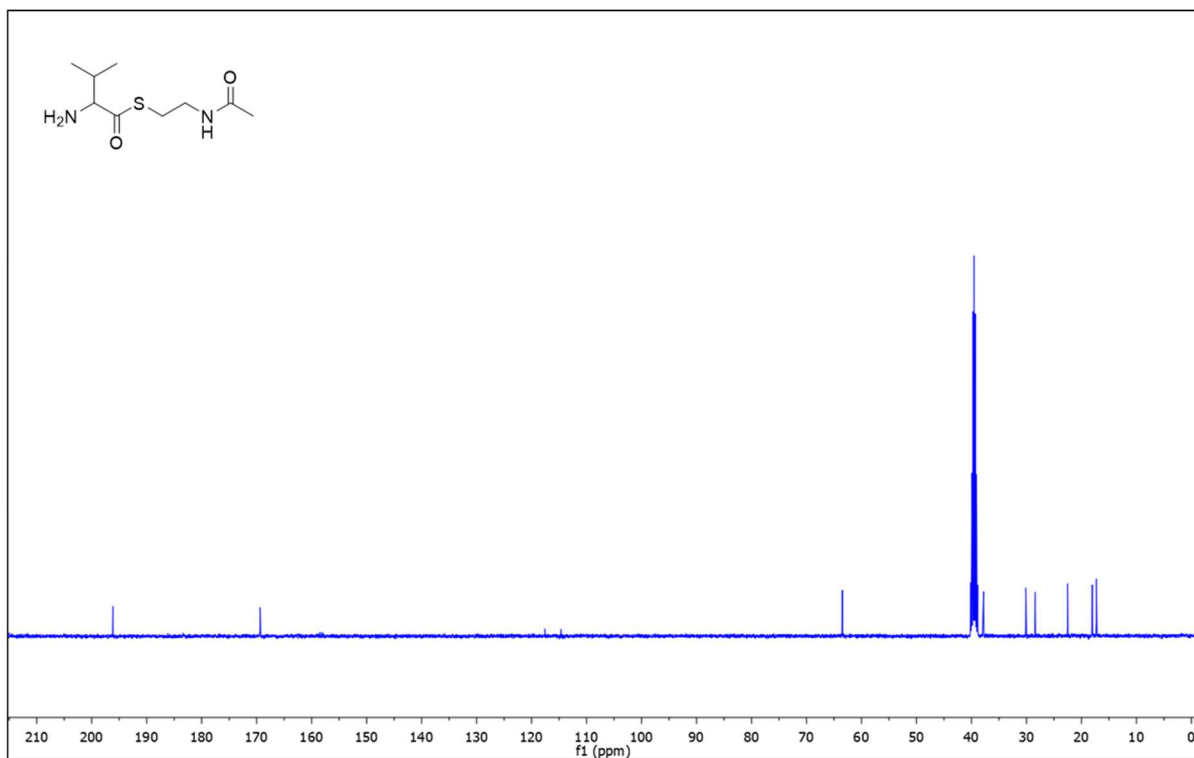
Figure S13. ^{13}C NMR spectrum of Val-SNAC (**9**) in DMSO-d_6 at 400 MHz.
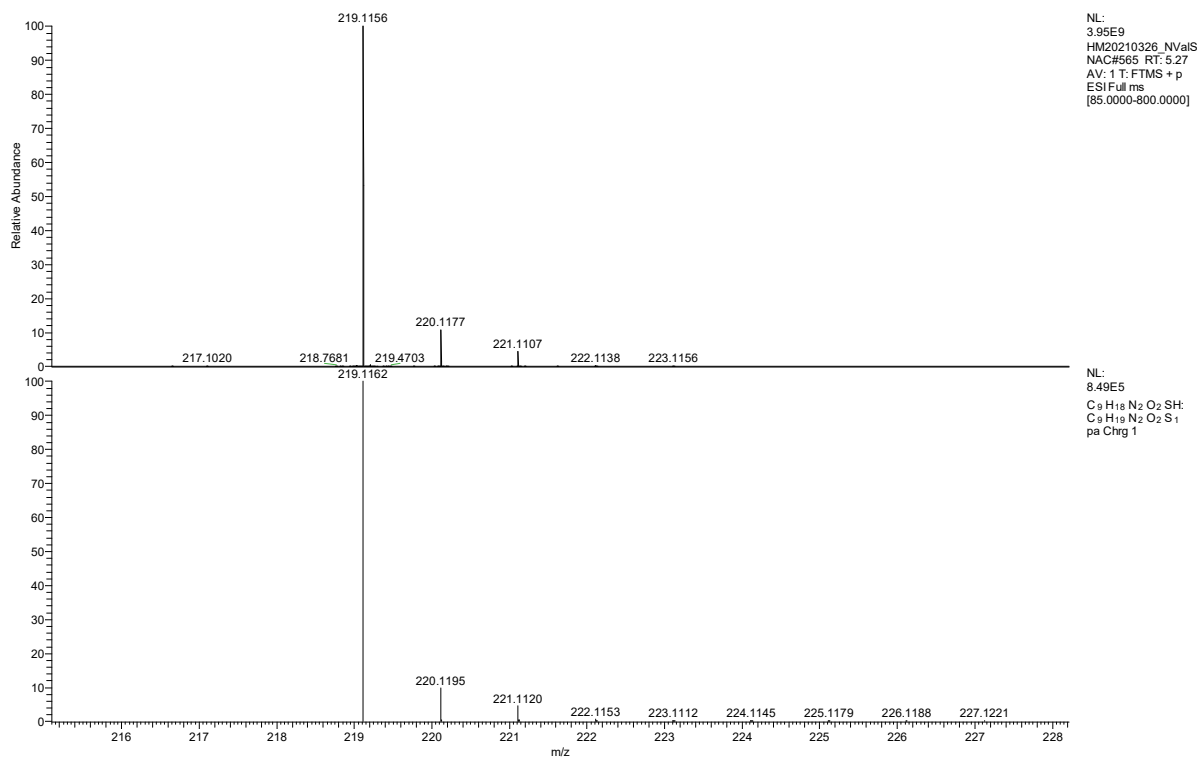
Figure S14. Top: mass spectrum for Nva-SNAC (**10**). Bottom: simulated mass spectrum.
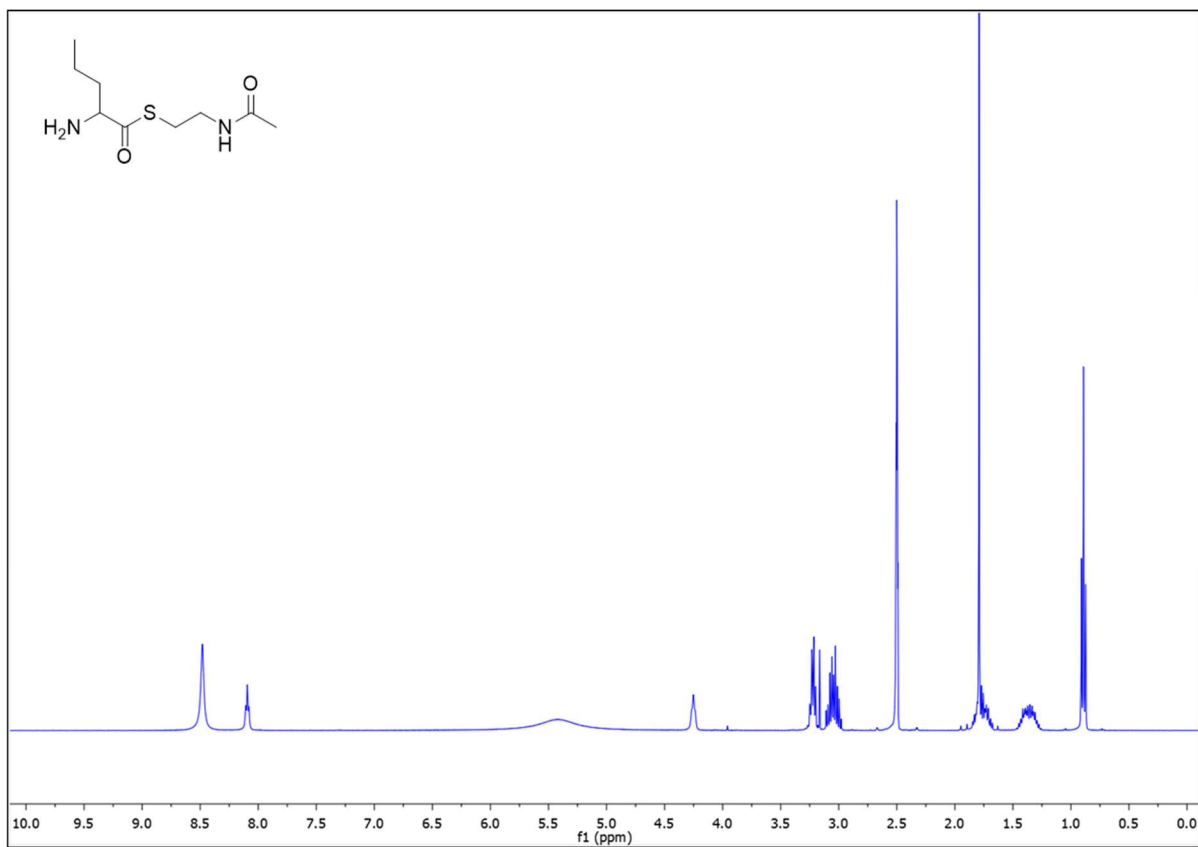
Figure S15. ^{1}H NMR spectrum of Nva-SNAC (**10**) in DMSO-d_6 at 400 MHz.
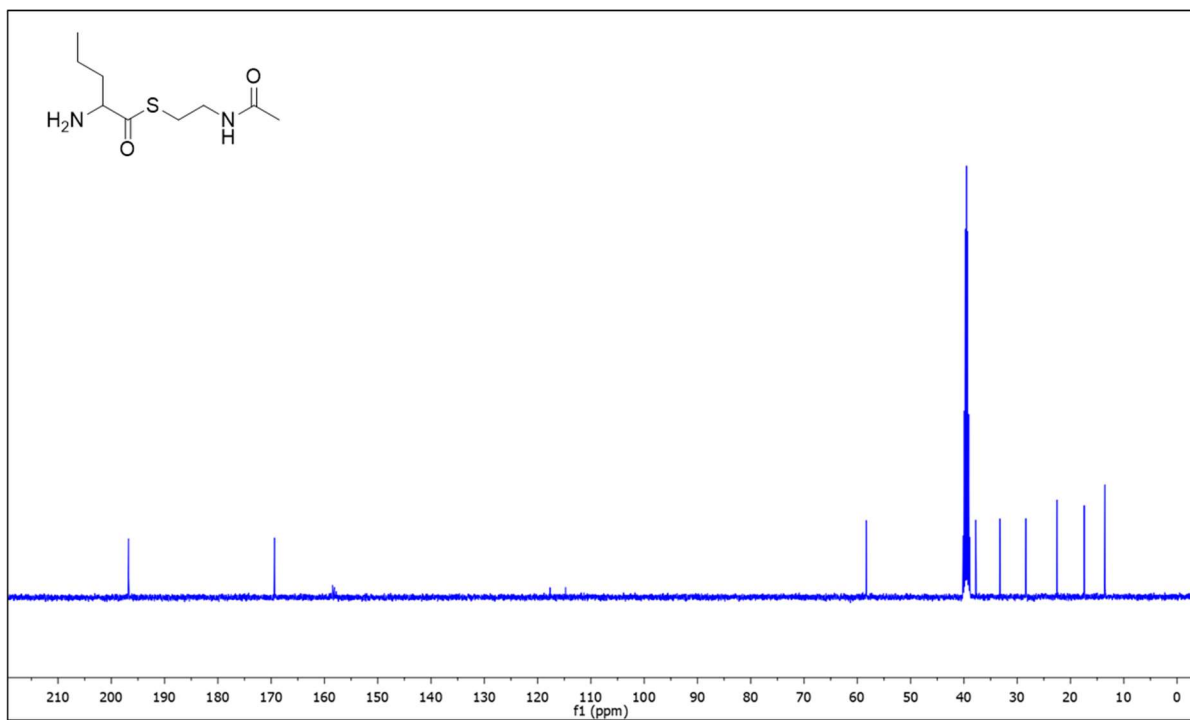
Figure S16. ^{13}C NMR spectrum of Nva-SNAC (**10**) in DMSO-d_6 at 400 MHz.
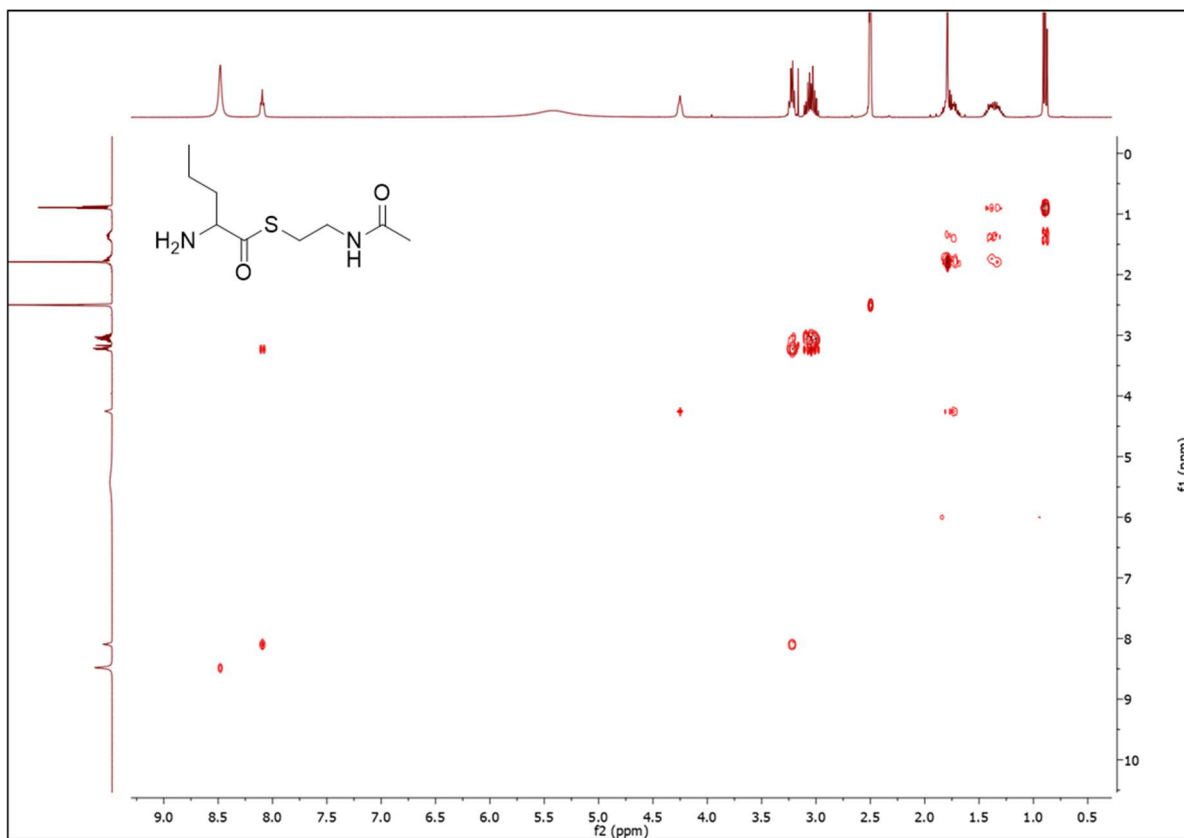
Figure S17. COSY spectrum of Nva-SNAC (**10**) in DMSO-d_6 at 400 MHz.

Figure S18. HSQC spectrum of Nva-SNAC (**10**) in DMSO-d_6 at 400 MHz.
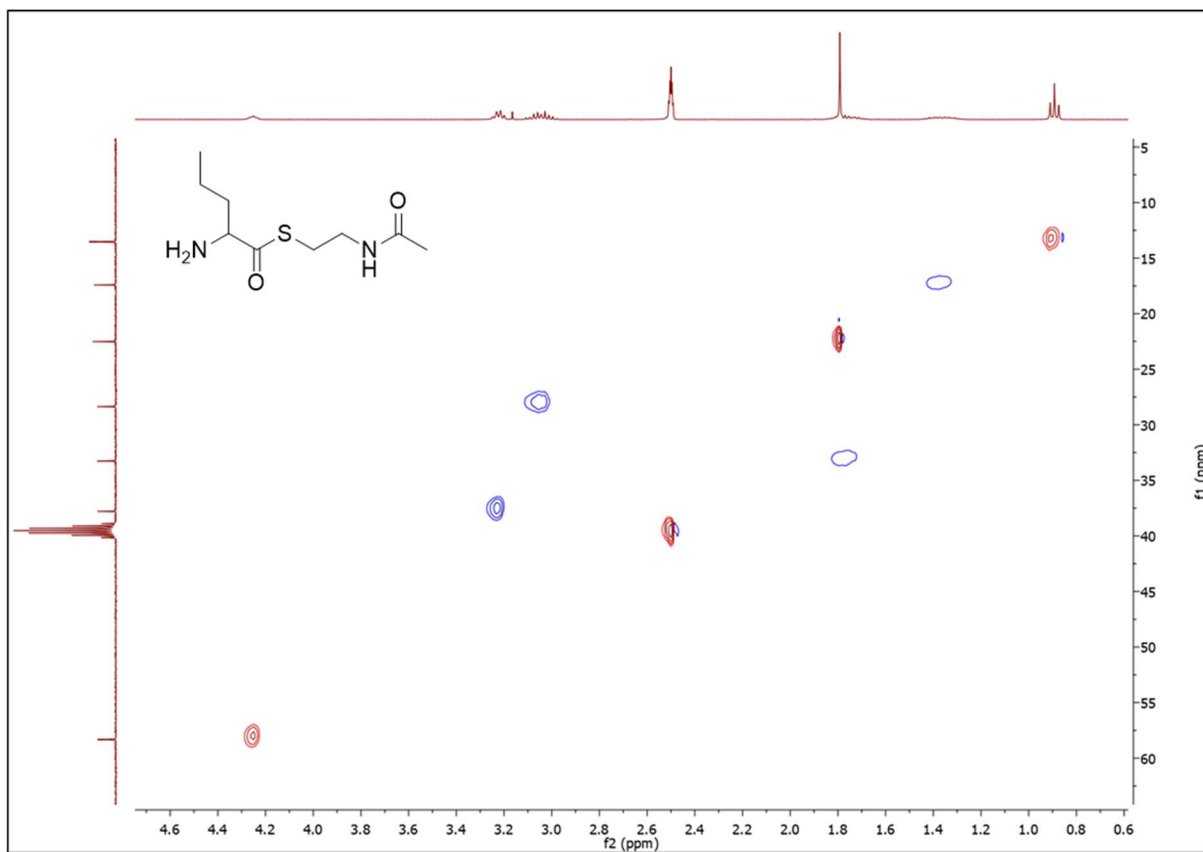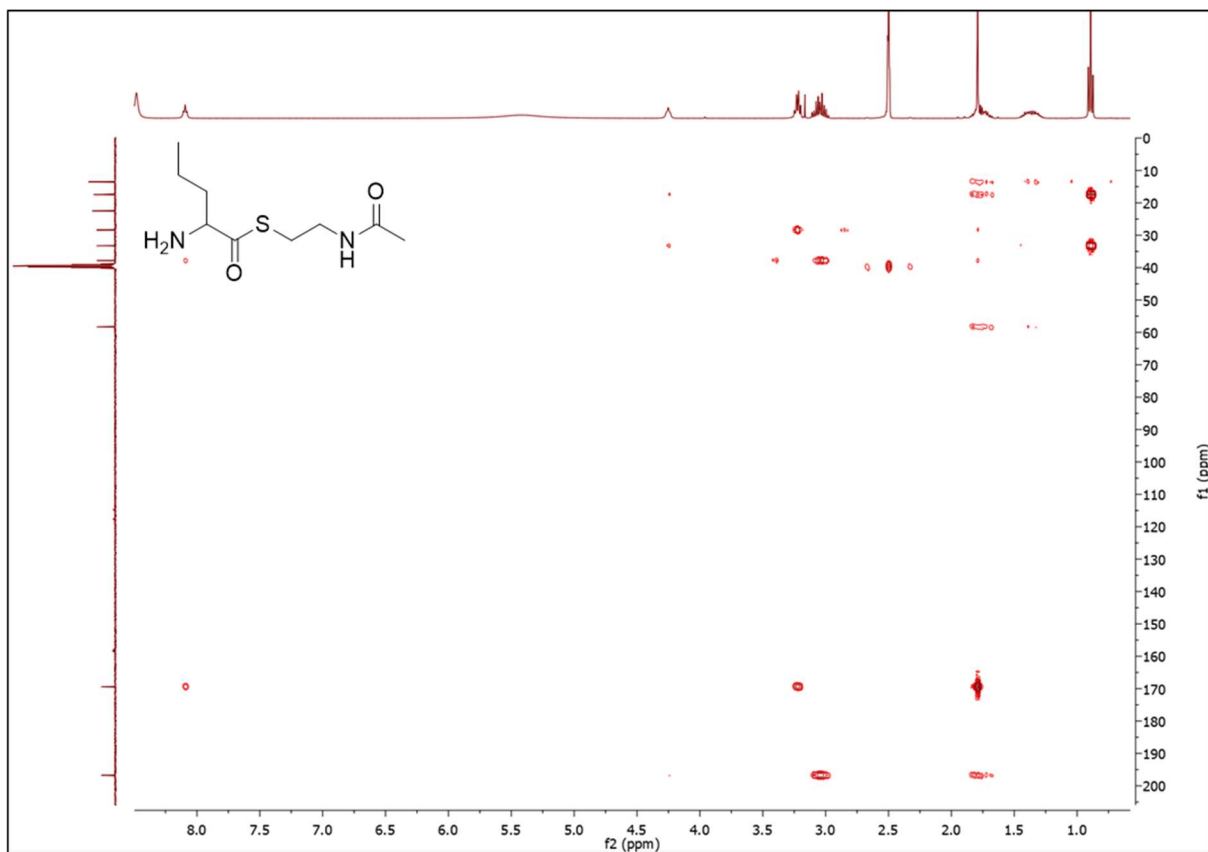
Figure S19. HMBC spectrum of Nva-SNAC (**10**) in DMSO-d_6 at 400 MHz.
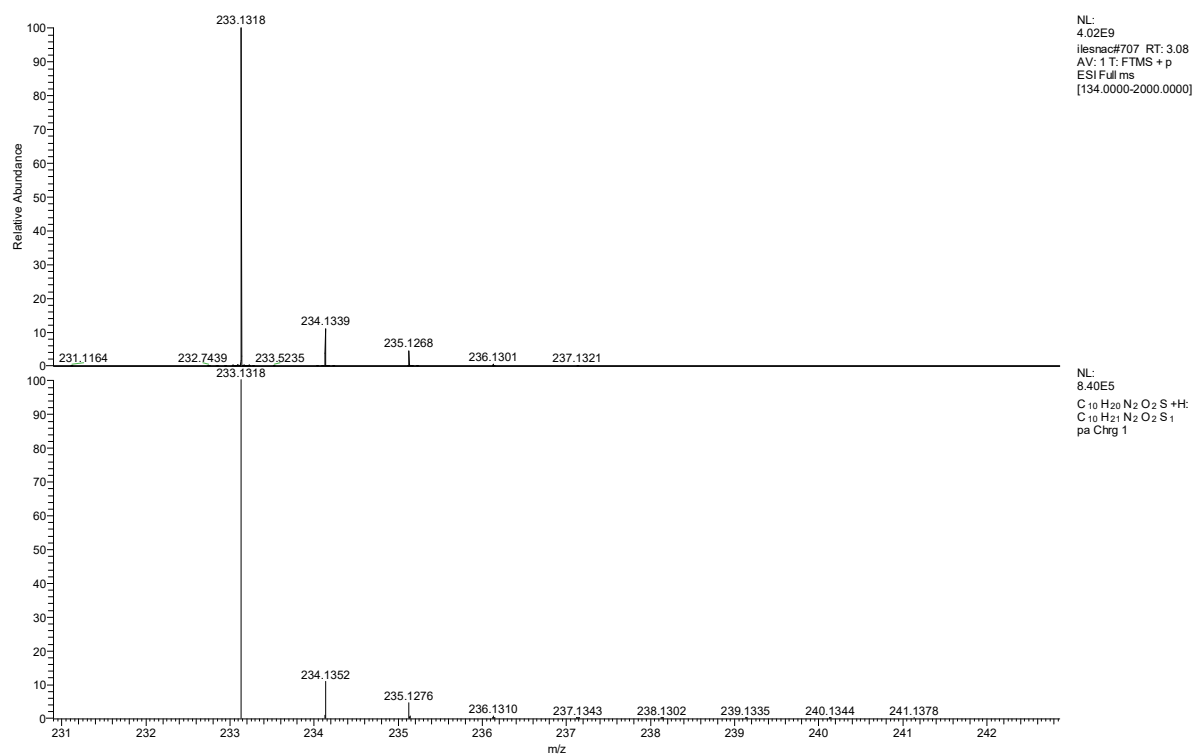
Figure S20. Top: mass spectrum for Ile-SNAC (**11**). Bottom: simulated mass spectrum.
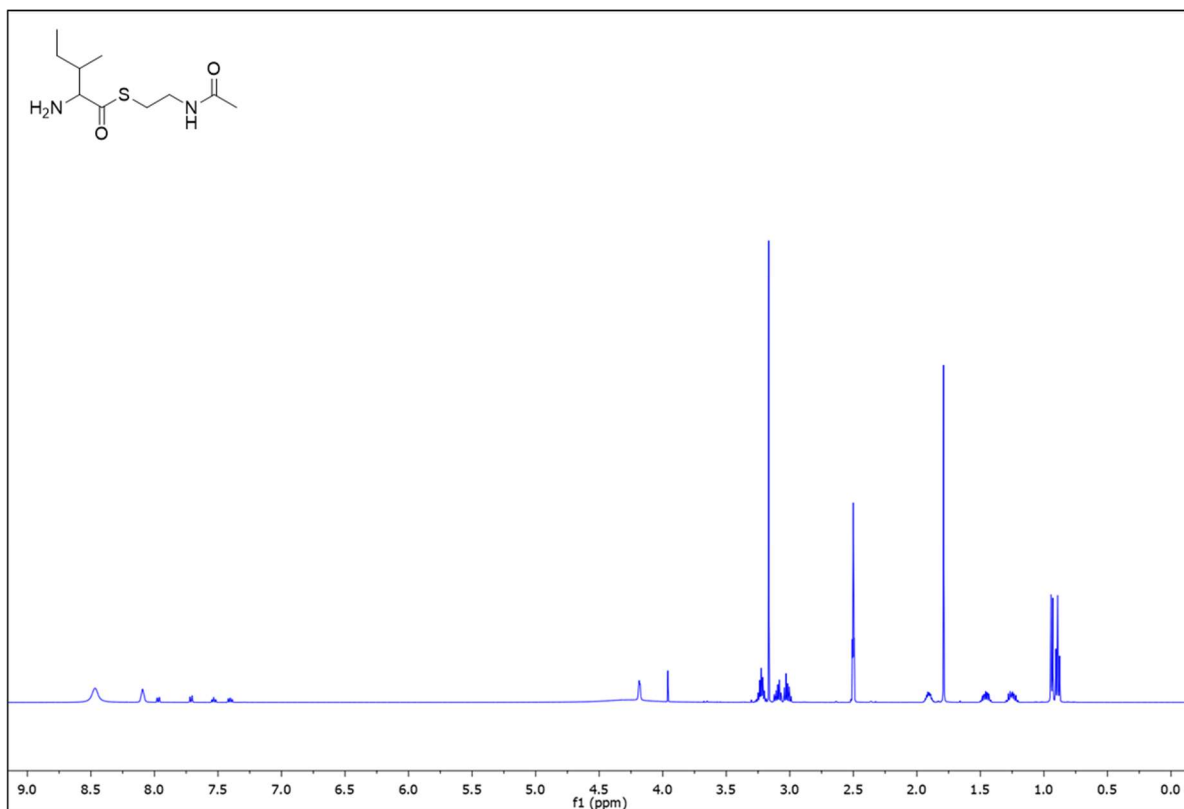
Figure S21. ^{1}H NMR spectrum of Ile-SNAC (**11**) in DMSO-d_6 at 500 MHz.
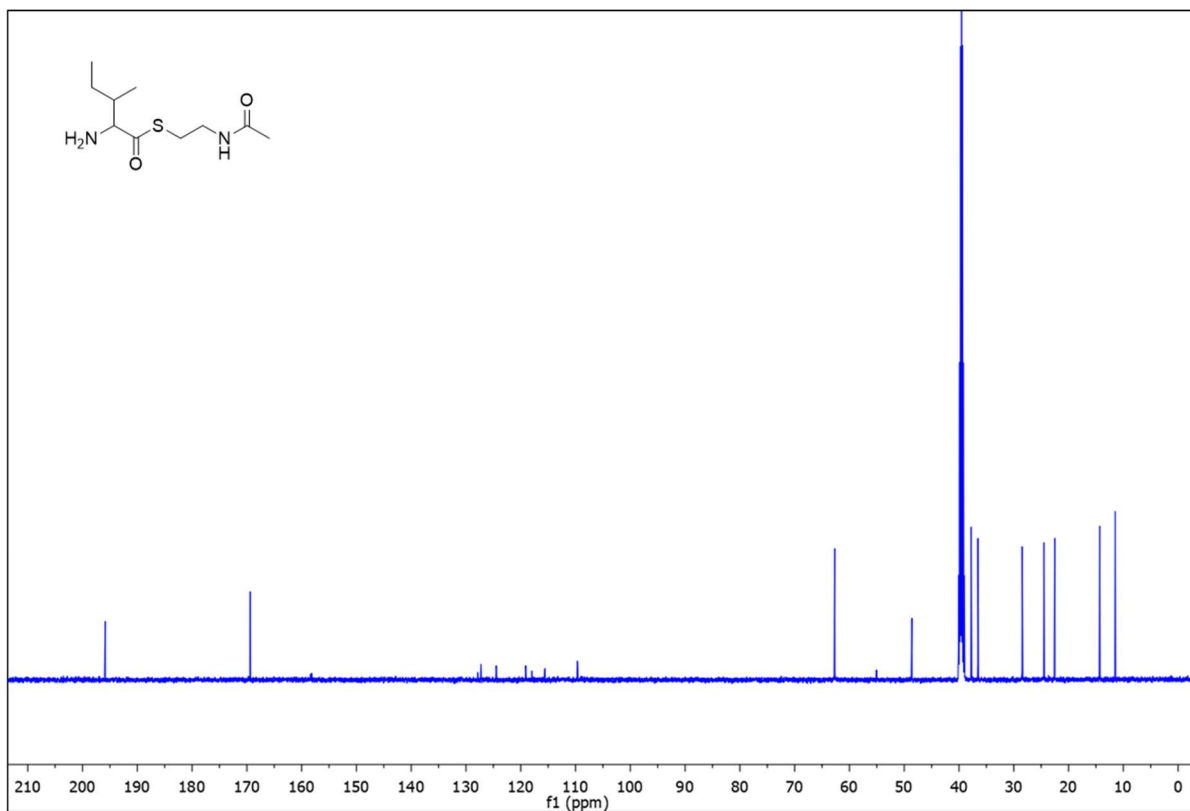
Figure S22. ^{13}C NMR spectrum of Ile-SNAC (**11**) in DMSO-d_6 at 500 MHz.
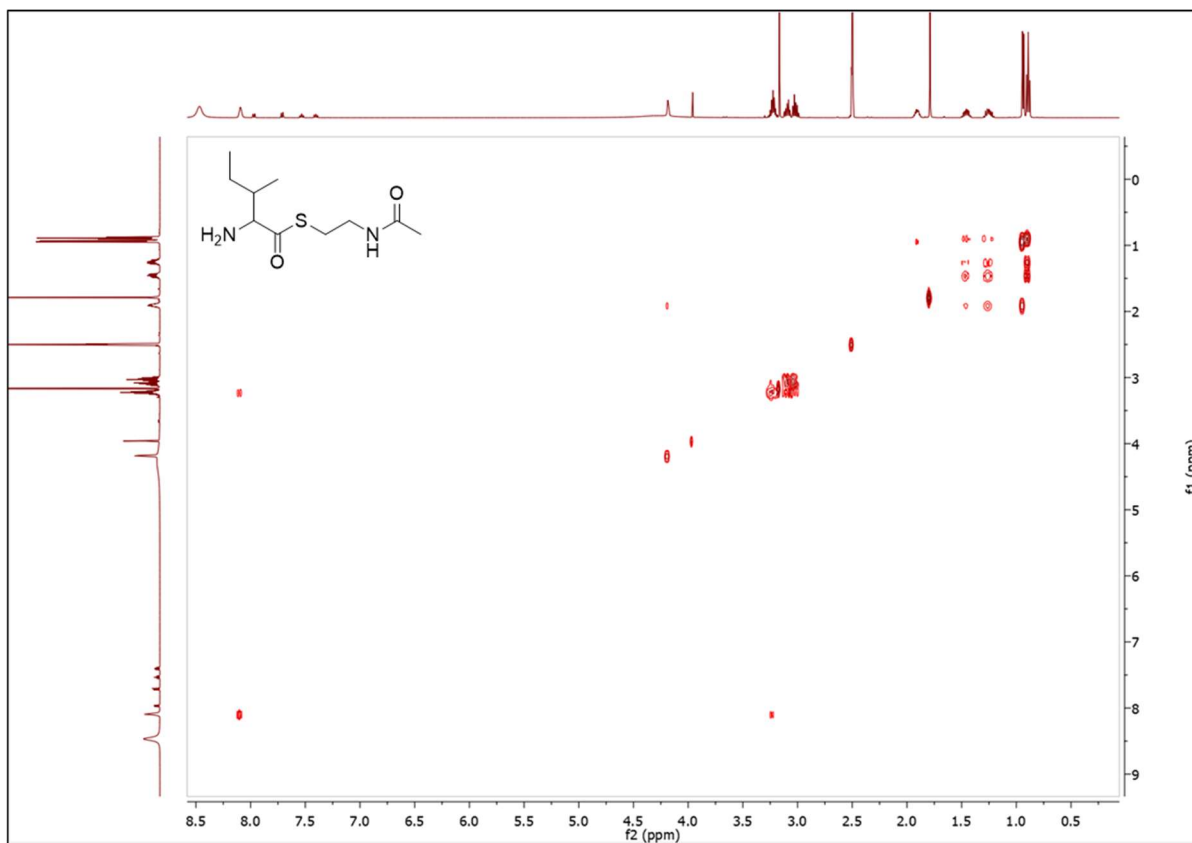
Figure S23. COSY spectrum of Ile-SNAC (**11**) in DMSO-*d₆* at 500 MHz.

Figure S24. HSQC spectrum of Ile-SNAC (**11**) in DMSO-d_6 at 500 MHz.
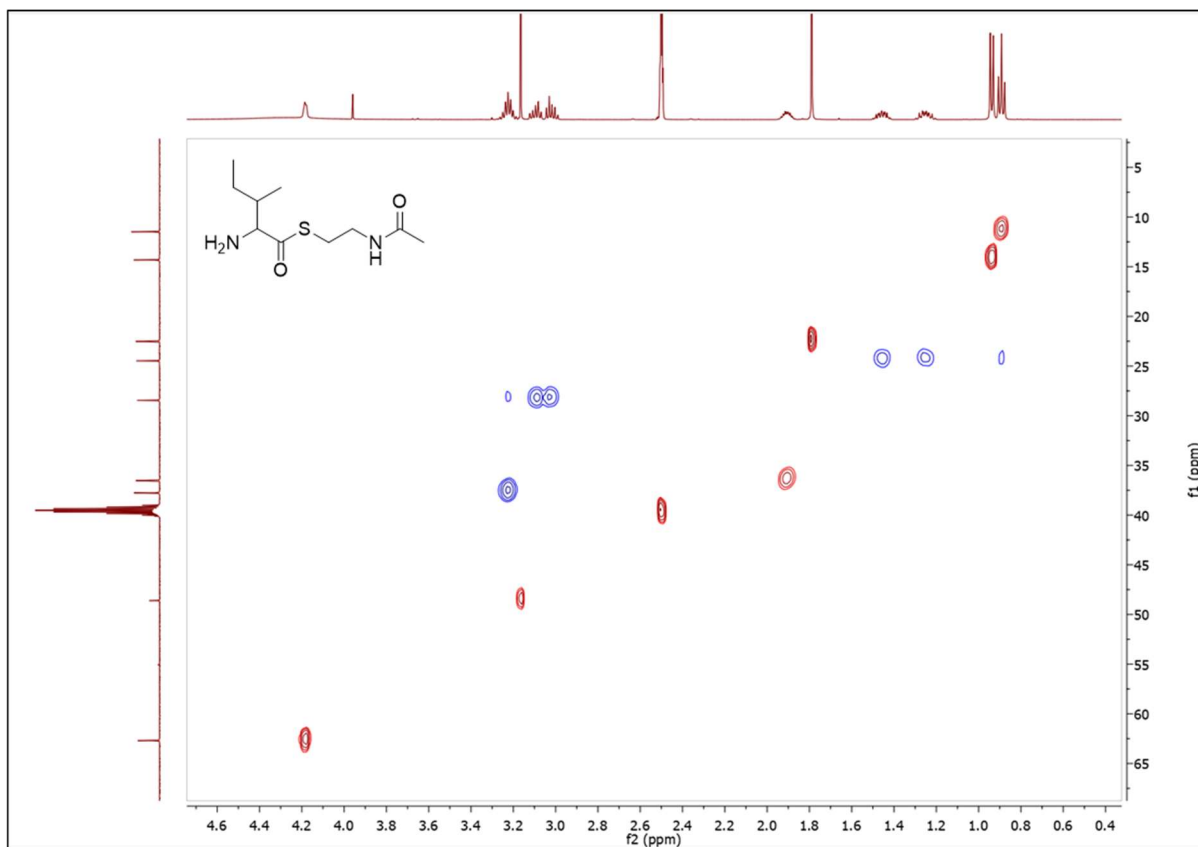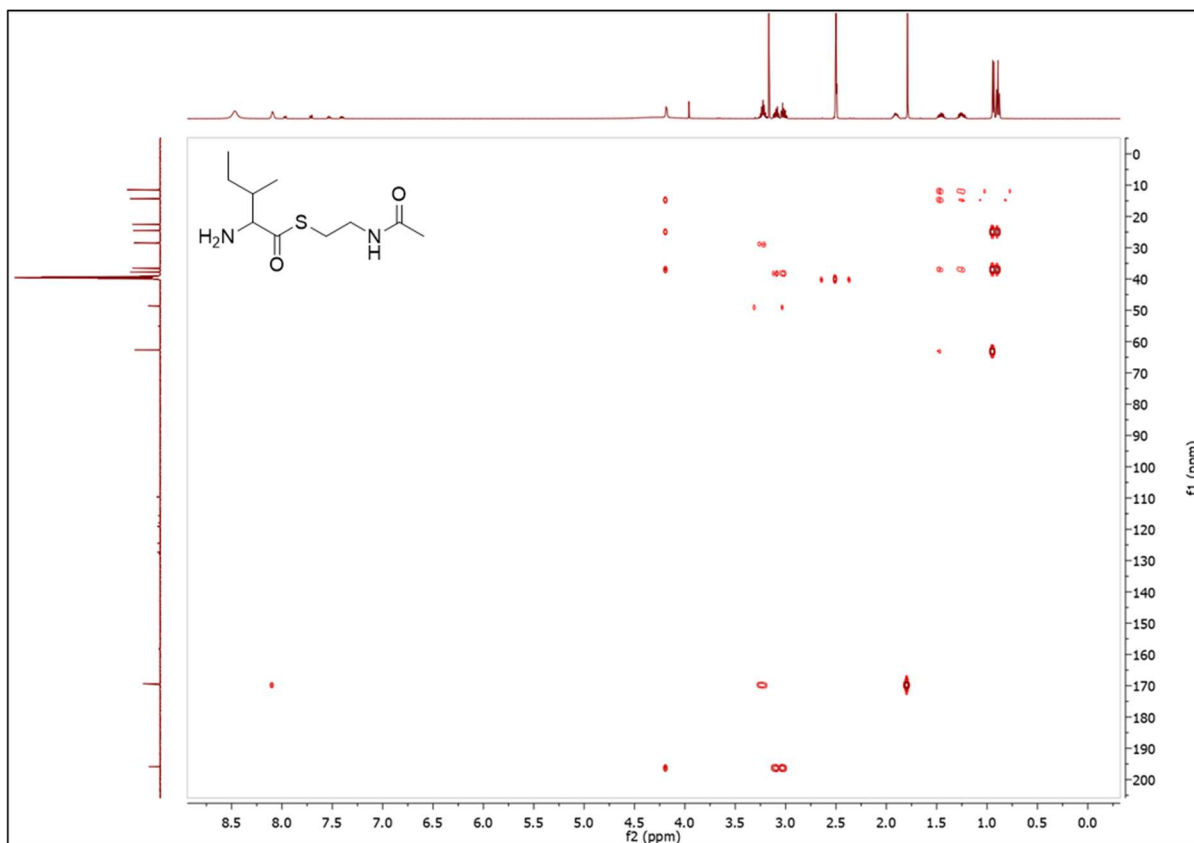
Figure S25. HMBC spectrum of Ile-SNAC (**11**) in DMSO-d_6 at 500 MHz.
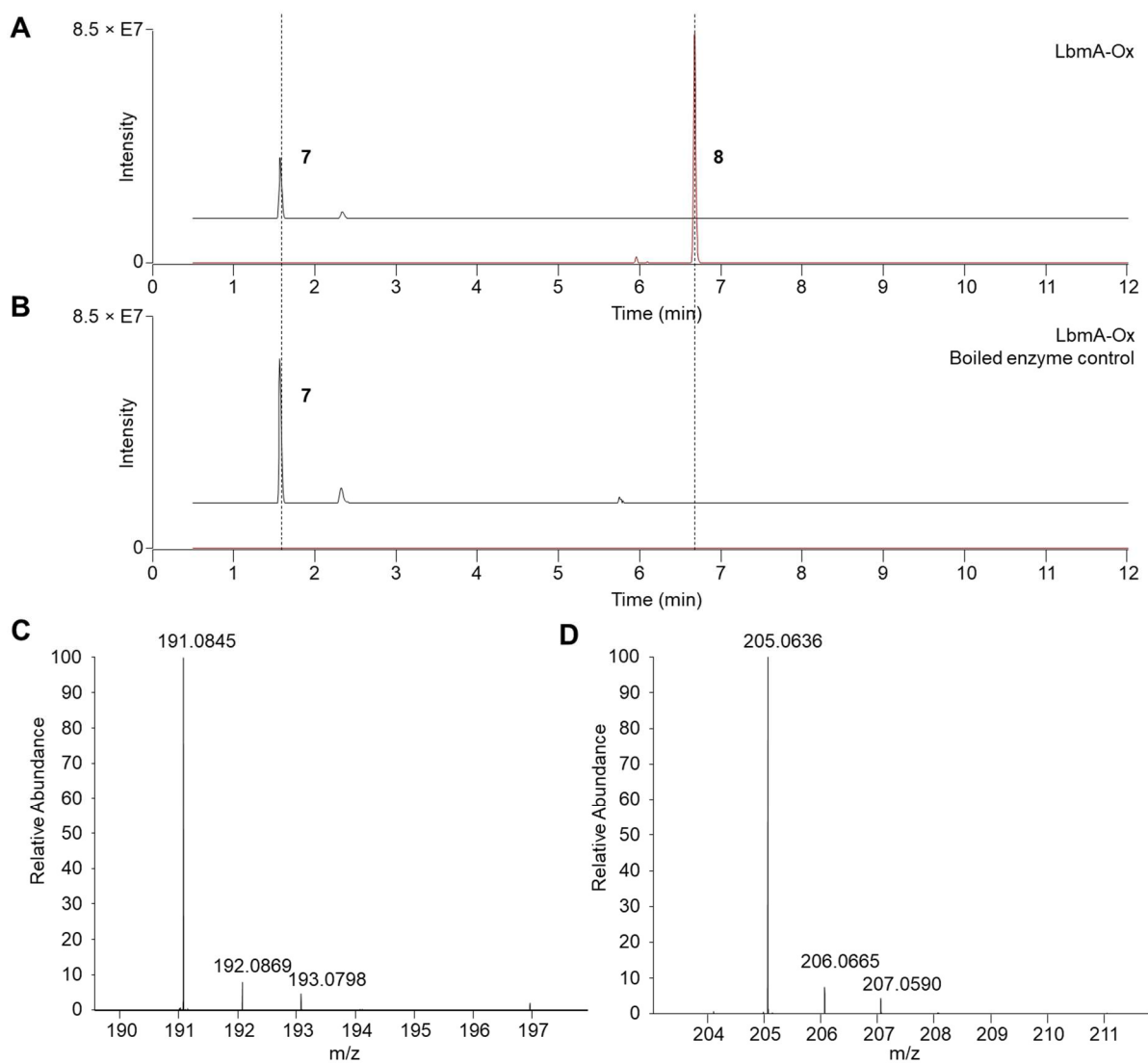
Figure S26. Extracted ion chromatogram (EIC) and mass spectra for *in vitro* LbmA-Ox assay mixtures using Ala-SNAC (**7**) as substrate. A) EIC for substrate **7** and product **8**. B) EIC for substrate **7** and product **8** for the control assay using boiled enzyme. C) Mass spectrum of **7** (calculated for $[M+H]^+$ 191.0848) at 1.58 min. D) Mass spectrum of **8** (calculated for $[M+H]^+$ 205.0641) at 6.68 min.

Figure S27. UHPLC-HRMS data of *in vitro* LbmA-Ox assay reaction with Ala-SNAC (**7**) as substrate. A) Total ion chromatogram. B) Extracted ion chromatogram for substrate **7**. C) Extracted ion chromatogram for product **8**. D) Filtered MS/MS spectrum of mass 205.1097 at 6.67 min. Black numbers indicate calculated *m/z* and red numbers indicate measured *m/z*. Possible fragment ions shown were deduced from predicted structures using the competitive fragmentation modeling for metabolite identification (CFM-ID 3.0) tool.[11]
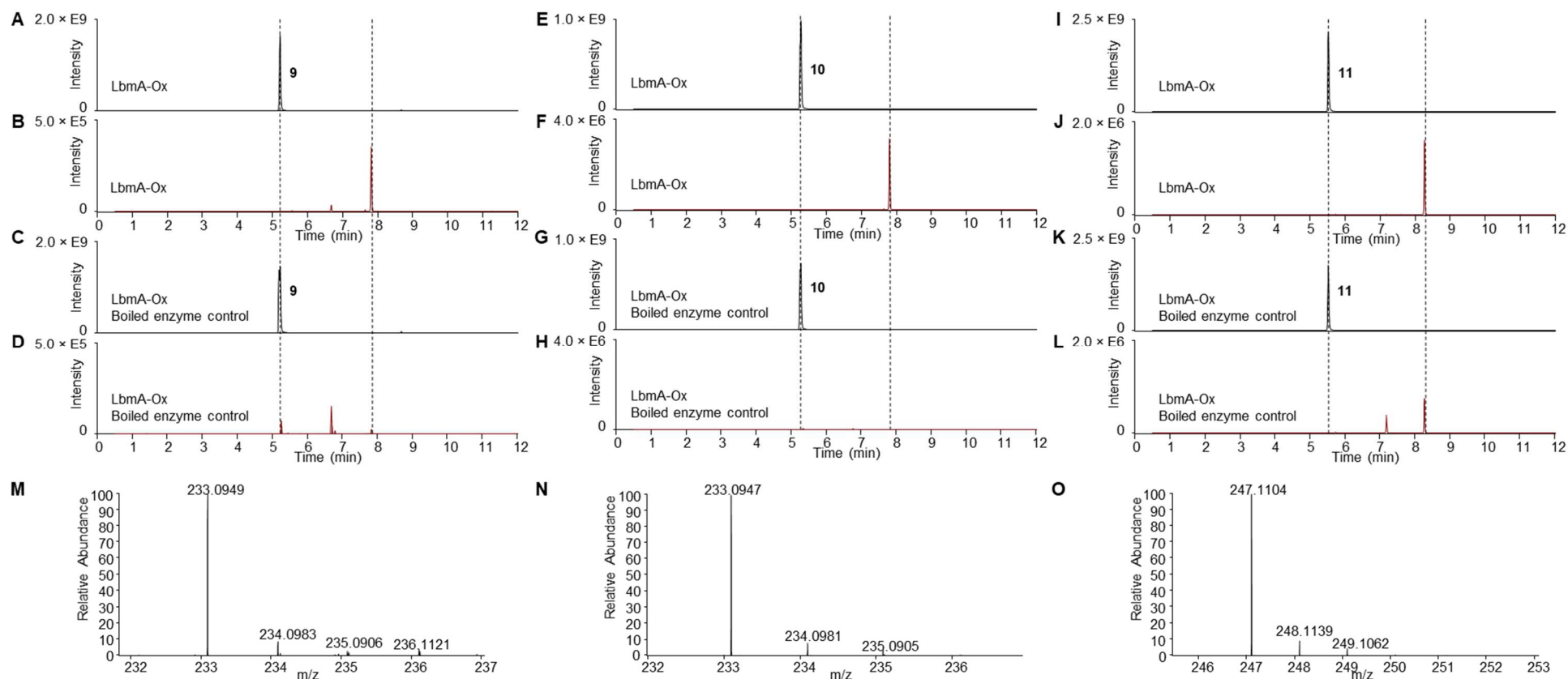
Figure S28. Extracted ion chromatograms (EIC) and mass spectra for assay mixtures using Val-SNAC (**9**), Nva-SNAC (**10**) or Ile-SNAC (**11**) as substrate. A) EIC for **9**. B) EIC for a putative product of **9**. C) EIC for **9** in the control assay. D) EIC for a putative product of **9** in the control assay. E) EIC for **10**. F) EIC for a putative product of **10**. G) EIC for **10** in the control assay. H) EIC for a putative product of **10** in the control assay. I) EIC for **11**. J) EIC for a putative product of **11**. K) EIC for **11** in the control assay. L) EIC for a putative product of **11** in the control assay. M) Mass spectrum for the putative product of **9** (calculated for [M+H]$^+$ 233.0954) at 7.82 min. N) Mass spectrum for the putative product of **10** (calculated for [M+H]$^+$ 233.0954) at 7.81 min. O) Mass spectrum for the putative product of **11** (calculated for [M+H]$^+$ 247.1111) at 8.28 min.
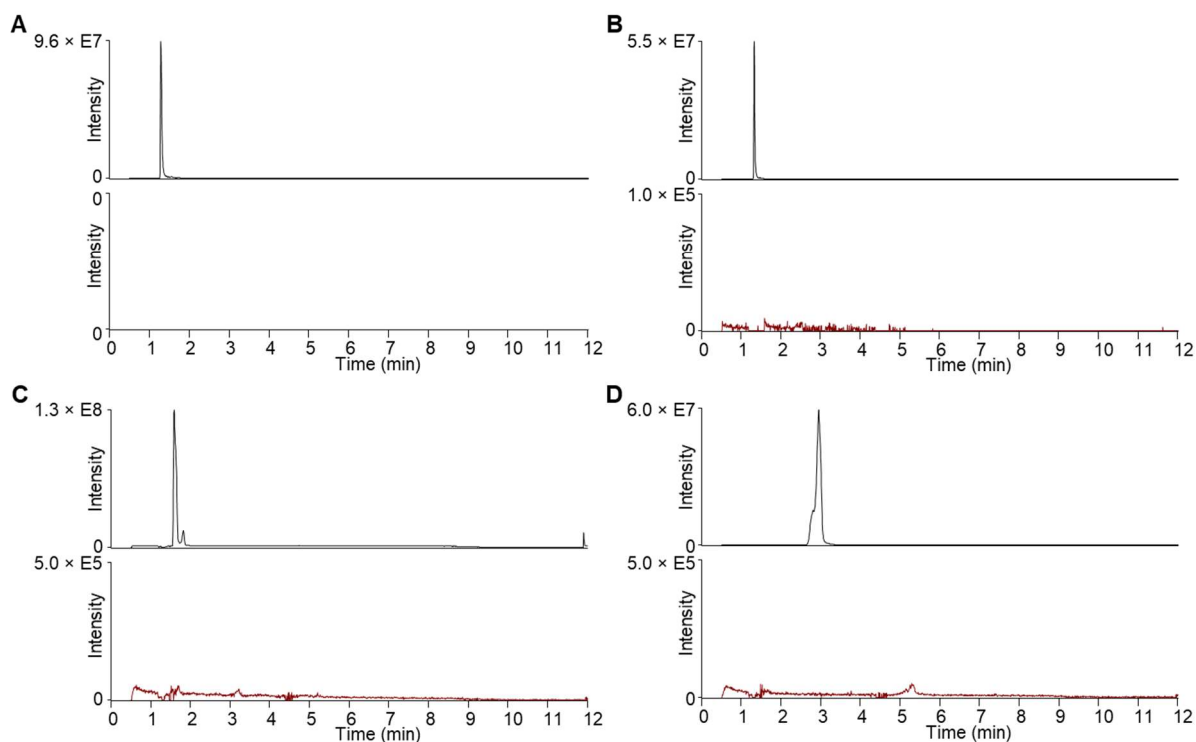
Figure S29. Extracted ion chromatogram (EIC) for *in vitro* LbmA-Ox assay mixtures using free amino acids as substrate. A) Upper trace: EIC for free glycine as substrate (calculated for [M+H]+ 76.0393). Lower trace: EIC for the product (calculated for [M+H]+ 90.0186). B) Upper trace: EIC for free alanine as substrate (calculated for [M+H]+ 90.0550). Lower trace: EIC for the product (calculated for [M+H]+ 104.0342). C) Upper trace: EIC for free valine as substrate (calculated for [M+H]+ 118.0863). Lower trace: EIC for the product (calculated for [M+H]+ 132.0655). D) Upper trace: EIC for free isoleucine as substrate (calculated for [M+H]+ 132.1019). Lower trace: EIC for the product (calculated for [M+H]+ 146.0812).
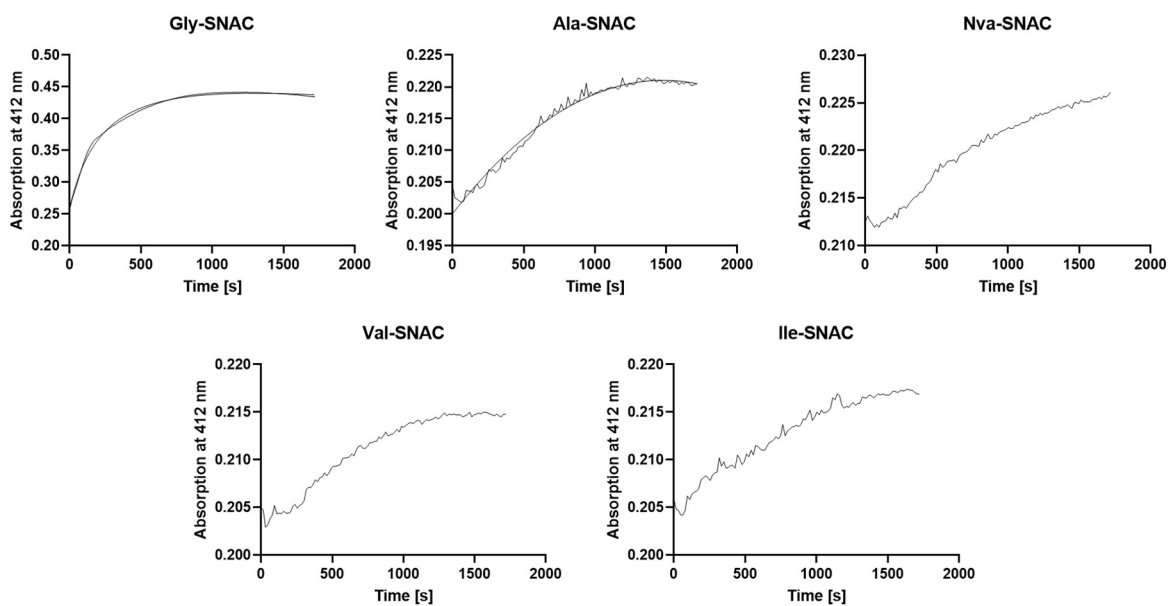
Figure S30. Ellman's assay of the different SNAC substrates. Absorption measurements at 412 nm over 2.5 h for the indicated substrates.
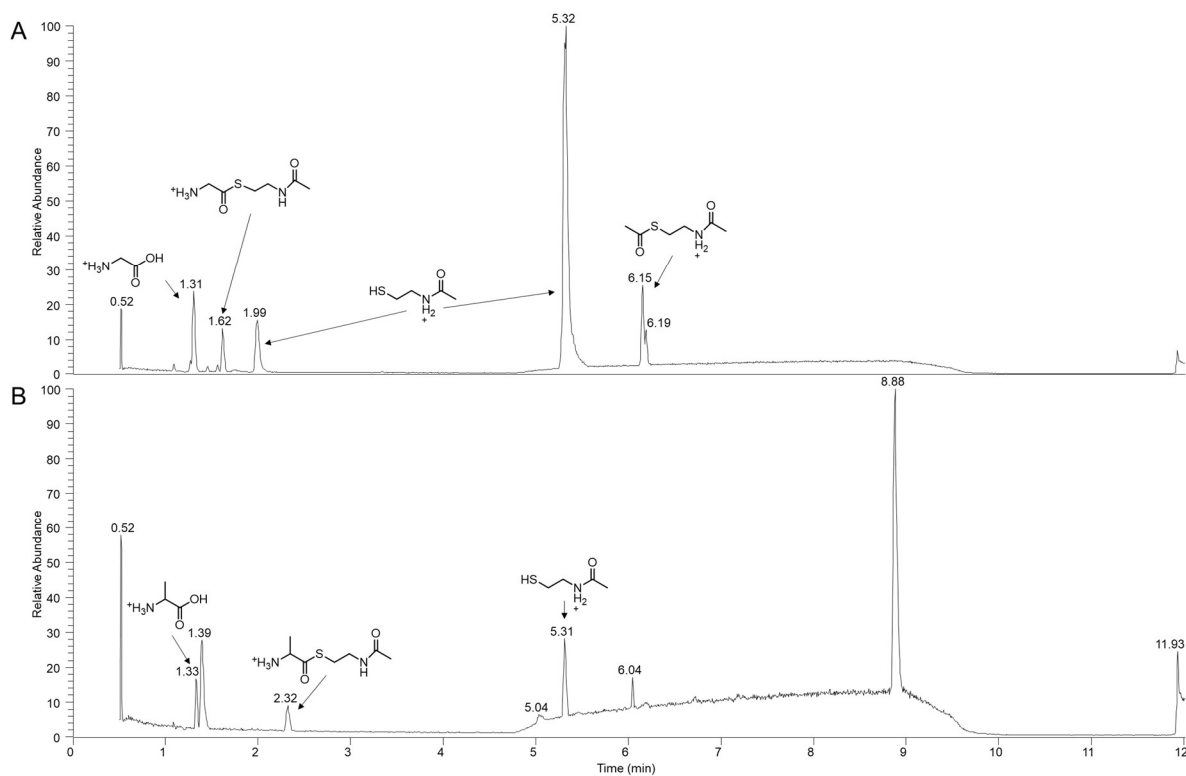
Figure S31. UHPLC-HRMS data of *in vitro* LbmA-Ox assay reactions using Gly-SNAC (**5**) or Ala-SNAC (**7**) that were hydrolyzed. A) Total ion chromatogram for an assay with **5** as substrate. B) Total ion chromatogram for an assay with **7** as substrate. The main peaks are assigned with putative hydrolysis products.
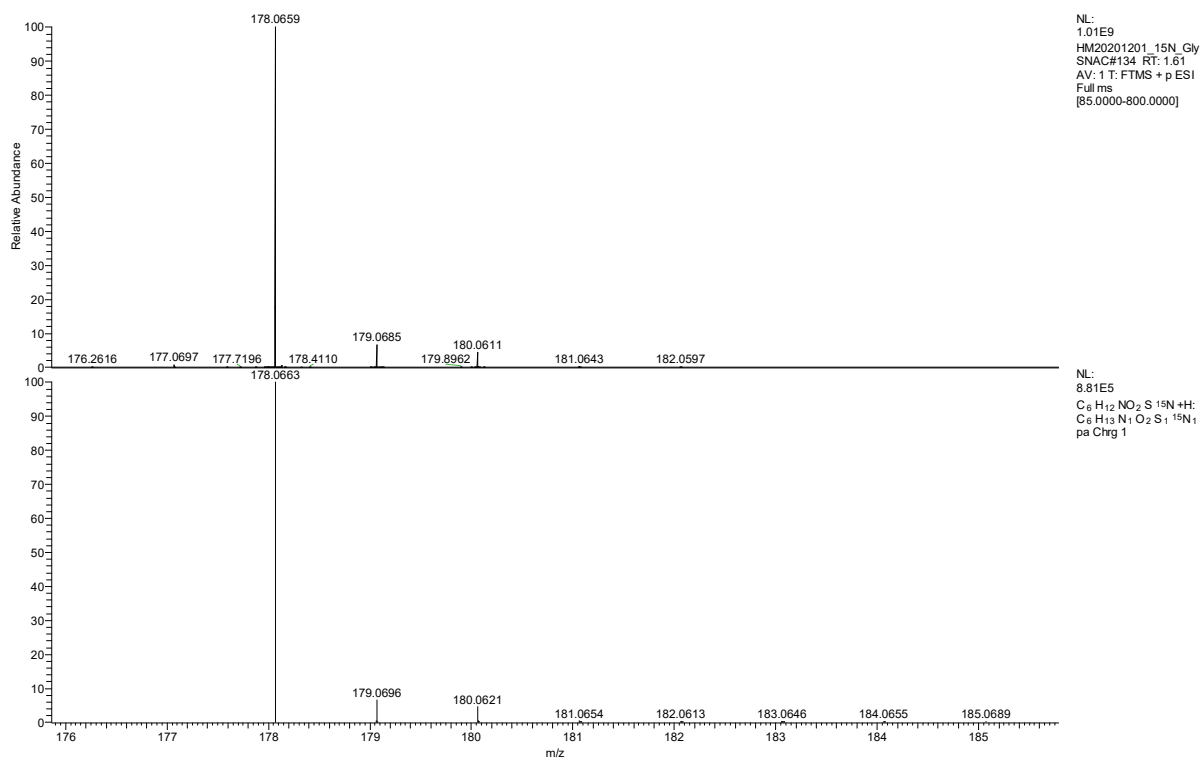
Figure S32. Top: mass spectrum for ^{15}N-Gly-SNAC (**12**). Bottom: simulated mass spectrum.
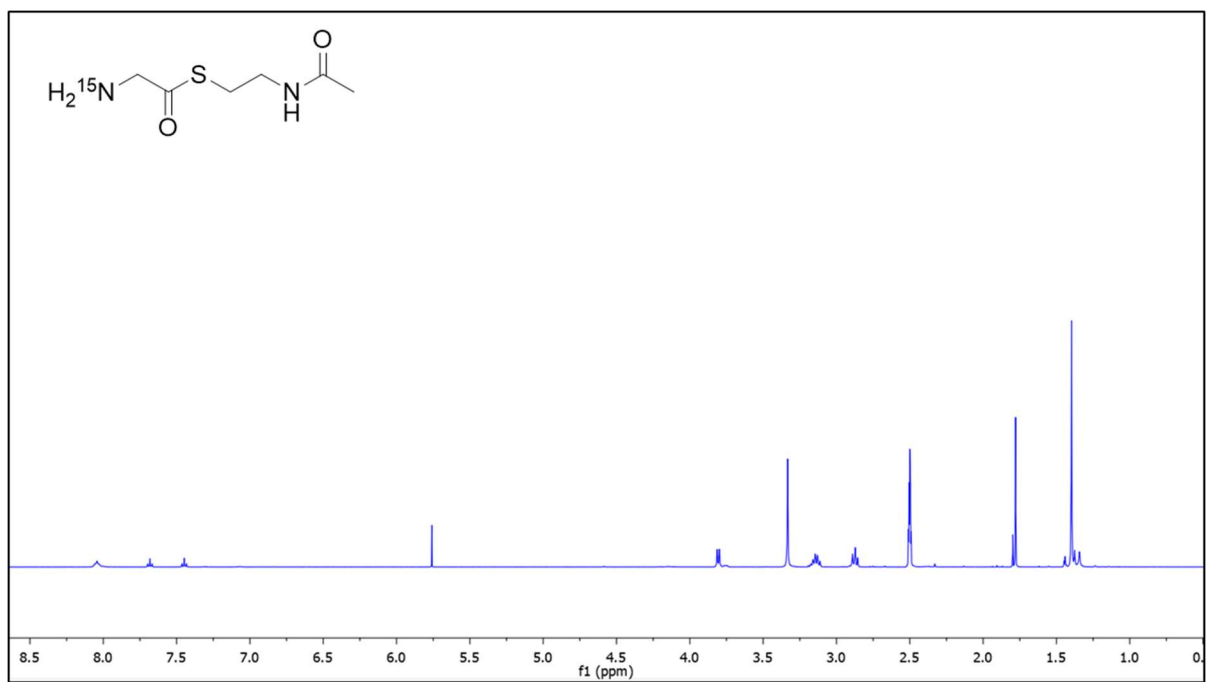
Figure S33. ^{1}H NMR spectrum of ^{15}N-Gly-SNAC (**12**) in DMSO-d_6 at 400 MHz.
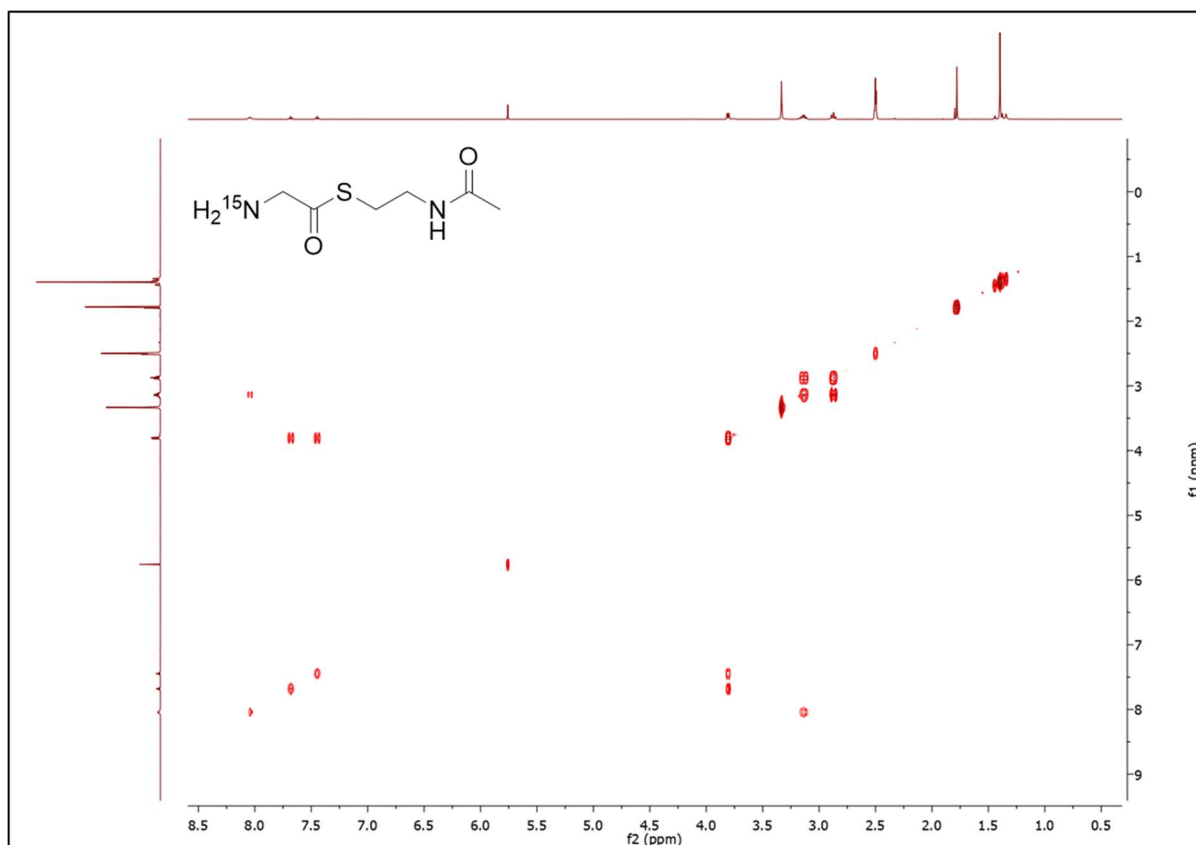
Figure S34. COSY NMR spectrum of ^{15}N-Gly-SNAC (**12**) in DMSO-d_6 at 400 MHz.
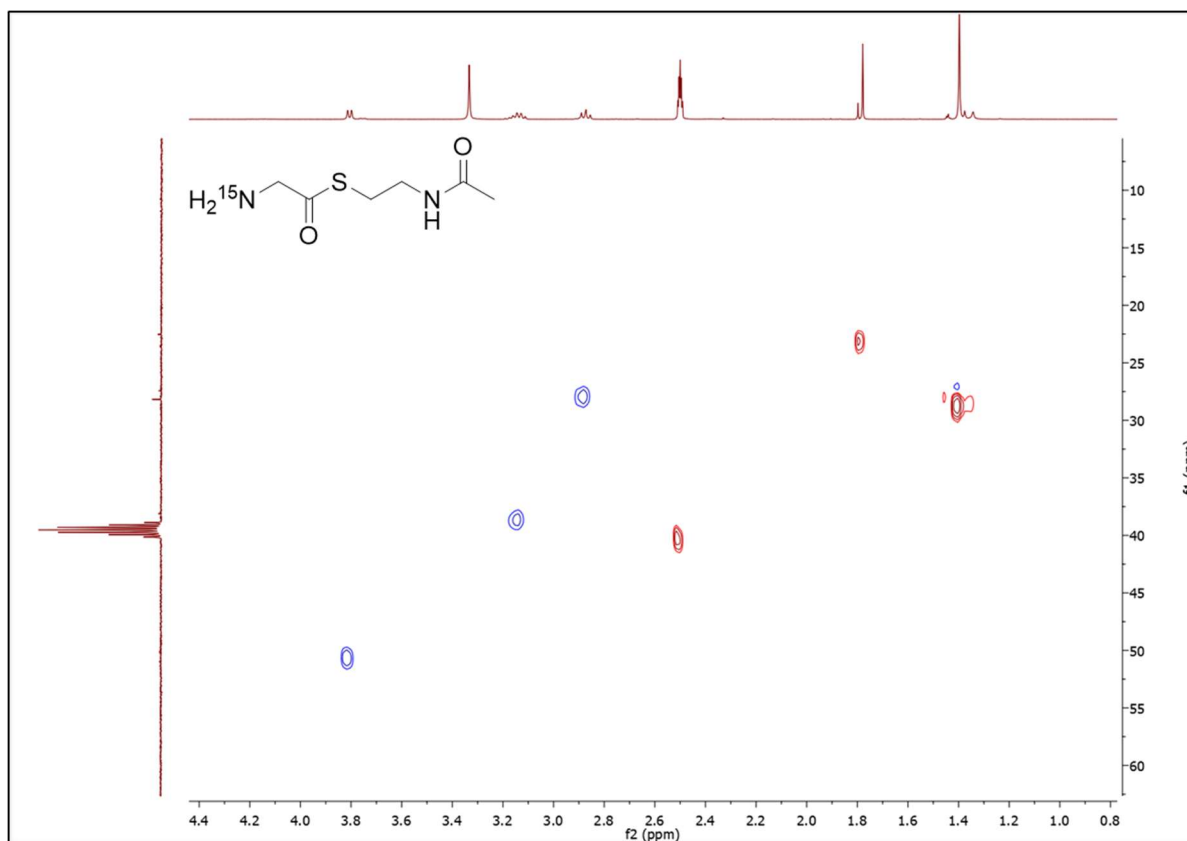
Figure S35. HSQC NMR spectrum of ^{15}N-Gly-SNAC (**12**) in DMSO-d_6 at 400 MHz.
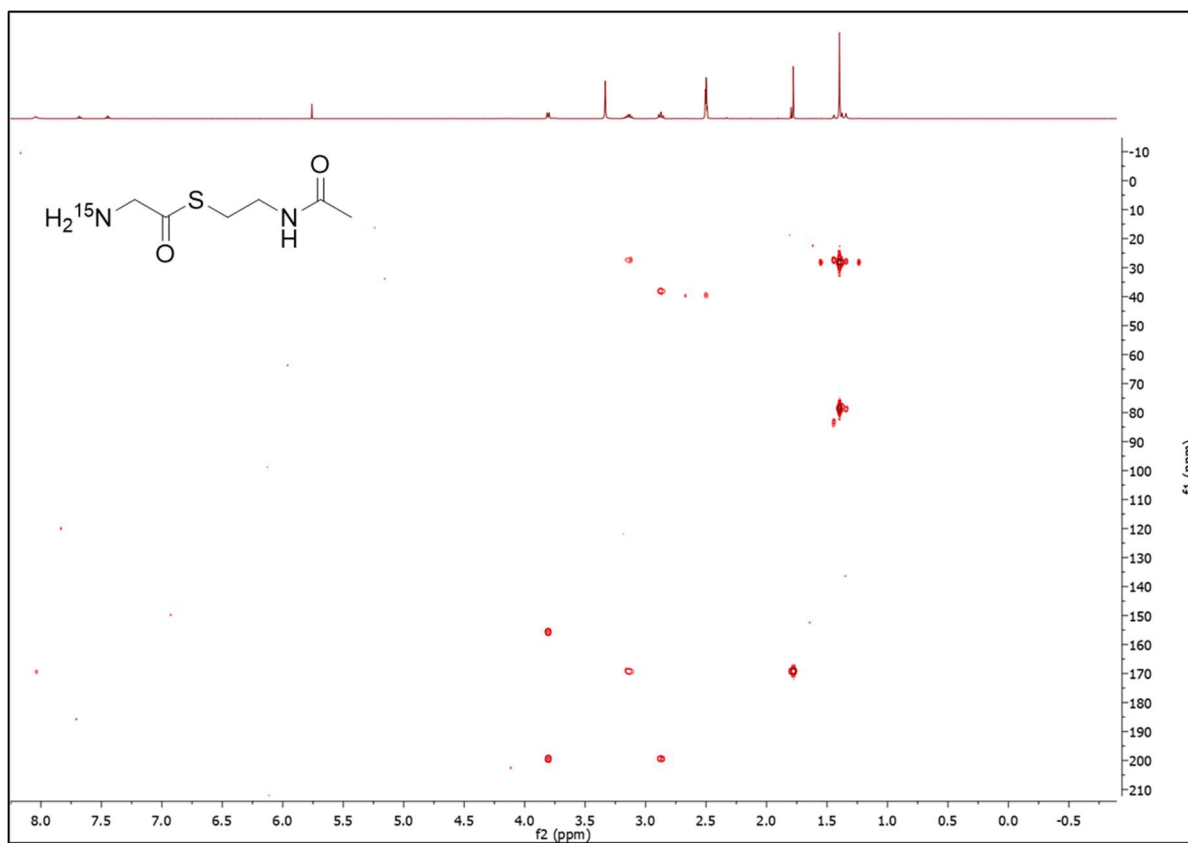
Figure S36. HMBC NMR spectrum of ^{15}N-Gly-SNAC (**12**) in DMSO-d_6 at 400 MHz.
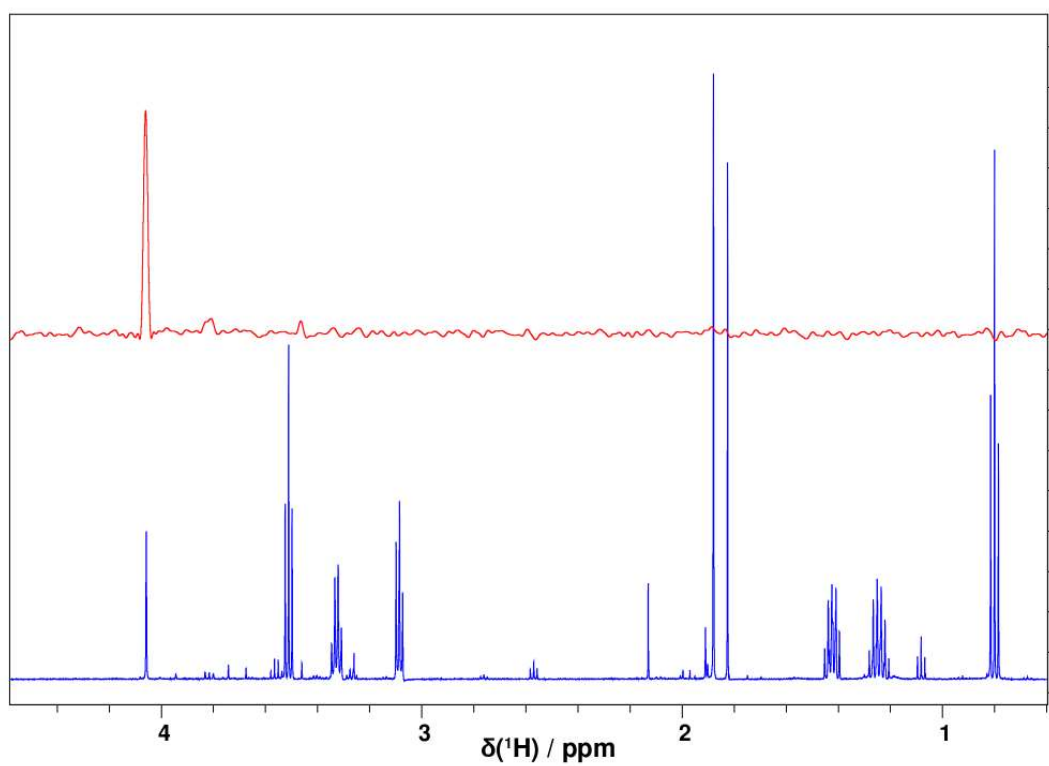
Figure S37. The upper trace in red shows the ^{15}N edited spectrum of ^{15}N-Gly-SNAC (**12**), the lower trace in blue shows the 1D spectrum of the same sample. The Hα signal is observable at 4.06 ppm. The Hα proton is weakly coupled to ^{15}N with a coupling constant of about 1 Hz.
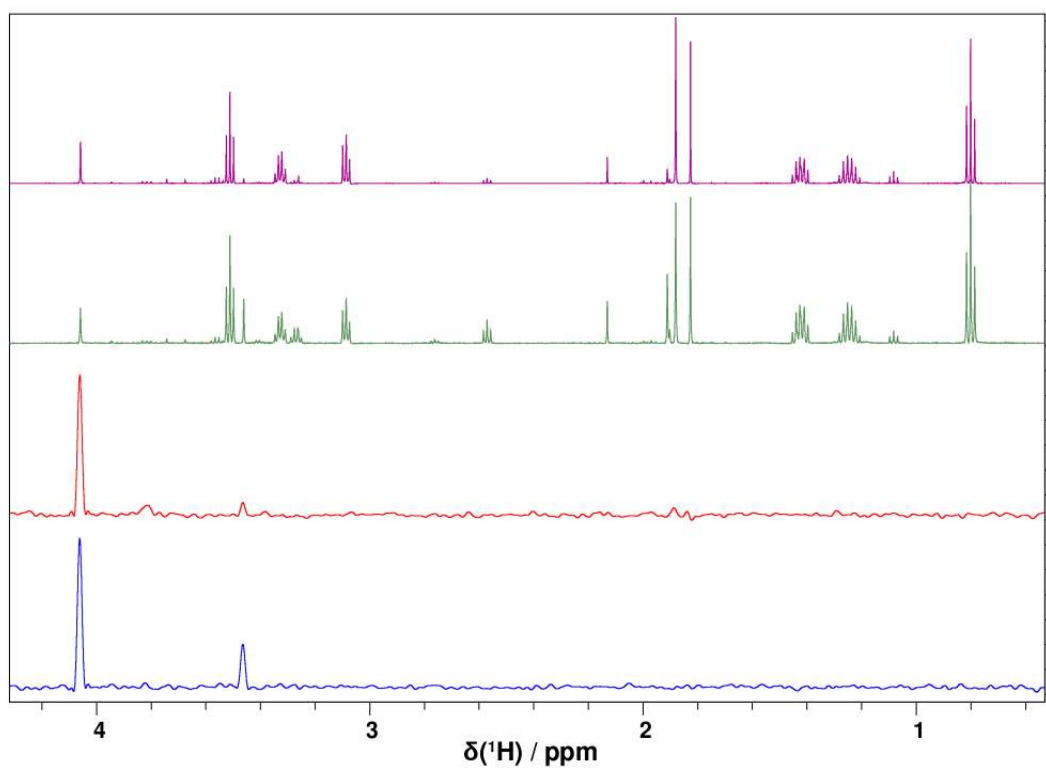
Figure S38. The lower trace in red shows the ^{15}N edited spectrum of ^{15}N-Gly-SNAC (**12**) at time 0 of an assay using LbmA-Ox directly observed in an NMR tube. The spectrum after 6 hours is shown in blue. The upper trace in pink shows the 1D spectrum of the same sample at time 0 and in green after 6 hours. The Hα signal is observable at 4.06 ppm. An additional signal appears at 3.46 ppm indicating degradation of **12** over time.
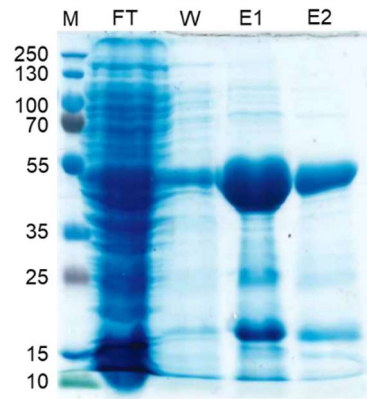
Figure S39. 12% SDS-PAGE gel of His$_6$-SUMO-tagged LbmA-MT post Ni-NTA affinity purification. Expected molecular mass: 91.3 kDa. M: protein ladder, FT: flowthrough, W: wash, E1: elution 1, E2: elution 2.
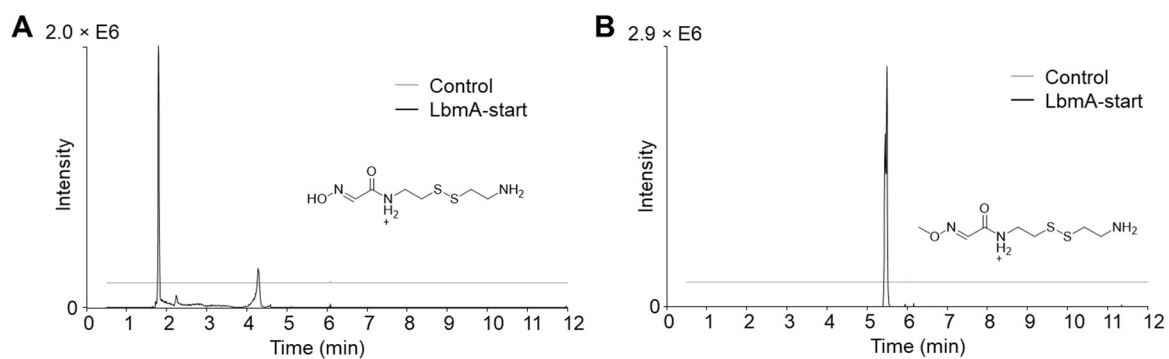
Figure S40. UHPLC-HRMS data of the *in vivo* assay using an empty plasmid as negative control. A) Extracted ion chromatogram (EIC) of **13** (calc. for [*M*+H]+ as 224.0522). B) Extracted ion chromatogram (EIC) of **14** (calc. for [*M*+H]+ as 238.0678).
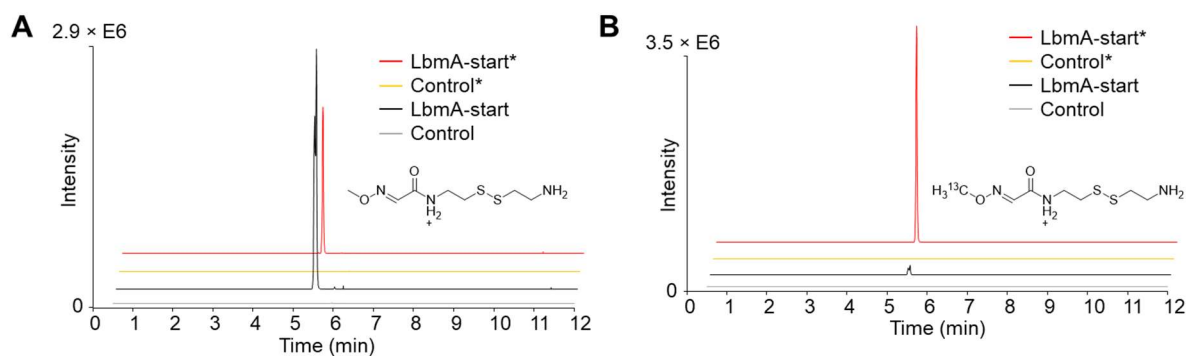
Figure S41. UHPLC-HRMS data of the *in vivo* assay using medium enriched with ^{13}C-methyl-methionine. A) Extracted ion chromatogram (EIC) of **14** (calc. for [*M*+H]$^+$ as 238.0678). *indicates growth in labelled medium, LbmA-start: assay using the starting modules of LbmA, control: assay using an empty vector control. B) Extracted ion chromatogram (EIC) of **14** that incorporated ^{13}C (calc. for [*M*+H]$^+$ as 239.0712). *indicates growth in labelled medium.

References

[1]   G. R. Fulmer, A. J. M. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw, K. I. Goldberg, *Organometallics* **2010**, *29*, 2176-2179.

[2]   R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792-1797.

[3]   M. N. Price, P. S. Dehal, A. P. Arkin, *Plos One* **2010**, *5*.

[4]   V. I. Morariu, B. V. Srinivasan, V. C. Raykar, R. Duraiswami, L. S. Davis, in *The Conference and Workshop on Neural Information Processing Systems*, **2008**, pp. 1113-1120.

[5]   D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343-345.

[6]   G. W. Heberlig, J. T. C. Brown, R. D. Simard, M. Wirz, W. Zhang, M. Wang, L. I. Susser, M. E. Horsman, C. N. Boddy, *Org. Biomol. Chem.* **2018**, *16*, 5771-5779.

[7]   S. W. Haynes, B. D. Ames, X. Gao, Y. Tang, C. T. Walsh, *Biochemistry* **2011**, *50*, 5668-5679.

[8]   V. Sklenar, M. Piotto, R. Leppik, V. Saudek, *J. Magn. Reson., Ser. A* **1993**, *102*, 241-245.

[9]   M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661-665.

[10]   M. Witanowski, *Pure Appl. Chem.* **1974**, *37*, 225-233.

[11]   Y. Djoumbou-Feunang, A. Pon, N. Karu, J. M. Zheng, C. Li, D. Arndt, M. Gautam, F. Allen, D. S. Wishart, *Metabolites* **2019**, *9*.

Chapter III

Manuscript in preparation

Thioesterase Exchanges in *trans*-Acyltransferase Polyketide Synthases

Hannah A. Minas[1§], Silke Reiter[2§], and Jörn Piel[1*]

[1]Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zürich, Vladimir-Prelog-Weg 4, 8093 Zürich (Switzerland).
E-mail: jpiel@ethz.ch

[2]Bioprocess Laboratory, Eidgenössische Technische Hochschule (ETH) Zürich, Mattenstrasse 26, 4058 Basel (Switzerland).

Author contributions

HAM, SR, and JP designed the research. SR designed the initial set of constructs. HAM performed expression experiments and HPLC-MS analysis. HAM and SR performed cloning, analyzed the data, and wrote the supporting information. HAM, SR, and JP wrote the manuscript.

Abstract

The discovery of new natural product scaffolds is essential for the development of antimicrobial agents to combat microbial resistance. Polyketides constitute a large class of natural products with a broad range of structural and functional diversity. They are produced by modular biosynthetic enzymes called type I polyketide synthases (PKS), which assemble the carbon backbone by stepwise incorporation of acyl building blocks. *Trans*-acyltransferase PKSs (*trans*-AT PKSs) are a subgroup of modular type I PKSs comprised of complex megaenzymes with interesting evolutionary mechanisms. Extensive horizontal gene transfer in these systems inspired us to test the possibility of engineering artificial *trans*-AT PKSs with new module combinations. Reprogrammed bacterial assembly lines have potential applications in the environmentally benign production of natural product analogs and drug scaffolds. One basic requirement of "designer" PKSs is the successful offloading of fully extended, non-native intermediates by a terminal thioester-cleaving domain. In this study, we assess the compatibility of various offloading domains with intermediates of the bacillaene *trans*-AT PKS. We generated several PKS hybrids with exchanged offloading modules and assessed the impact of different fusion points. The results indicate that the key for successful offloading of foreign polyketides is the inclusion of one acyl carrier protein native to each pathway. Additionally, the incorporation of linker regions is detrimental to PKS function. Furthermore, we identified a *trans*-acting condensation domain capable of offloading the final product bacillaene, indicating its potential as a new terminal domain in *trans*-AT PKS engineering efforts. These results provide a foundation from which to build the first generation of engineered *trans*-AT PKS clusters.

Introduction

Almost two thirds of all small-molecule drugs in the clinic are in some way related to natural products, either as the original compound or as a drug-lead that was further derivatized.[1] Of these, polyketides are amongst the most diverse in structure and function.[2] Some polyketide families are synthesized by megaenzymatic assembly lines termed polyketide synthases (PKSs), which produce chemical complexity from small CoA-activated acyl building blocks such as malonyl-CoA.[2] They are comprised of modules that in turn contain a variable array of enzymatic domains. In the subtype of *trans*-acyltransferase PKSs (*trans*-AT PKSs), a minimal module contains an acyl carrier protein (ACP) which shuttles the intermediate that remains attached to the PKS as thioesters from domain to domain and a ketosynthase domain (KS) responsible for successive polyketide chain elongation. The building blocks for chain extension are provided by a free-standing acyl transferase (AT) that acts *in trans*. Further optional modifications are carried out by ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains that are jointly capable of consecutively reducing the β-keto group to a fully saturated carbon-carbon bond.[2] Methyltransferases (MT) are another type of domain commonly observed in *trans*-AT PKSs. Ultimately, the polyketide chain is released from the PKS assembly line by cyclization or hydrolysis, typically but not always facilitated by a thioesterase domain (TE).[2-3]

Upon the discovery of PKSs, their modularity and collinear biosynthetic logic inspired researchers to engineer them for combinatorial biosynthesis.[4] These efforts have almost exclusively focused on textbook *cis*-AT PKSs - assembly lines containing an AT domain in every module - with mixed results, rarely yielding dramatically altered skeletons.[5] One potentially decisive difference between textbook PKSs and *trans*-AT PKSs is their mode of evolution. Due to the propensity of *trans*-AT PKSs to shuffle modules between different pathways naturally by horizontal gene transfer,[6] we hypothesized that these assembly lines would be more tolerant towards recombineering than *cis*-AT PKSs, which mainly evolve by vertical gene transfer and duplication of modules.[7]

A crucial step in engineered, hybrid assembly lines is the release of the final intermediate.[8] This step is typically performed by a TE, yet terminal condensation domains (Cs), which are commonly found in non-ribosomal peptide synthetases (NRPSs), have also been reported in *trans*-AT PKSs.[3a] The general mechanism of TEs relies on a conserved catalytic triad (Ser-His-Asp) that transfers the final intermediate from the last ACP onto the active site serine of the TE. Depending on the nucleophile that attacks the bound intermediate, the polyketide is released as a linear (hydrolysis) or macrocyclized (intramolecular cyclization) product (Figure 1A). Further, TEs can catalyze oligomerization during which the final polyketide is kept in the "waiting room" provided by the TE until a second intermediate is processed.[9] Based on structural insights in *cis*-AT PKSs, the TE forms a substrate channel to accommodate the final intermediate.[10] Changes in pH as well as structural components lead to conformational changes which in turn facilitate a preferred release mechanism. In the case of the *cis*-AT 6-deoxyerythronolide B synthase (DEBS) involved in erythromycin biosynthesis, substrate loading is controlled by hydrogen bond interactions[11] and macrolactonization is favored by shielding the active site pocket from water.[12] This way, intramolecular release is favored over hydrolysis.[9] While release of an unnatural compound was found to be a bottleneck in PKSs,[13] there does not seem to be a clear relation between TE sequences and their release mode. In general, TEs form distinct phylogenetic clades according to the type of biosynthetic assembly line they originate from: NRPSs, *cis*-AT PKSs, or *trans*-AT PKSs.[12b]

Given the modular evolution of *trans*-AT PKSs, we here studied the potential to exchange terminal modules in the archetypical bacillaene-producing *trans*-AT PKS (Figure 1B).[14] Due to the inherent instability of the compound, Moldenhauer et al. have previously attempted to produce truncated versions of this PKS to detect potentially more stable intermediates.[14a] When either the terminal TE domain was deleted or the TE was introduced downstream of module 6, successively elongated but stalled

biosynthetic intermediates were detected as hydrolyzed products by mass spectrometry.[14a] The ease with which the entire bacillaene route is observable in this way and the facile genetic manipulation of *Bacillus subtilis* (*B. subtilis*) prompted us to utilize it as a model system. In this work, we tested whether the levels of released final polyketide can be restored by fusing non-native terminal modules to the assembly line. Further, the tendency of stalled intermediates to be hydrolyzed can be exploited to pinpoint an impaired transfer of the intermediate onto the non-native TE or its functionality. Here we produce a range of terminal module swaps by introducing corresponding modules from seven *trans*-AT PKSs into the bacillaene *trans*-AT PKS. By testing different variants of the recombination site for best release of bacillaene, we have put in place the first set of rules for the generation of functional hybrid *trans*-AT PKS assembly lines. Jointly, these experiments provide the first insights into the feasibility of recombineering in *trans*-AT PKSs towards production of novel therapeutically relevant complex polyketides.
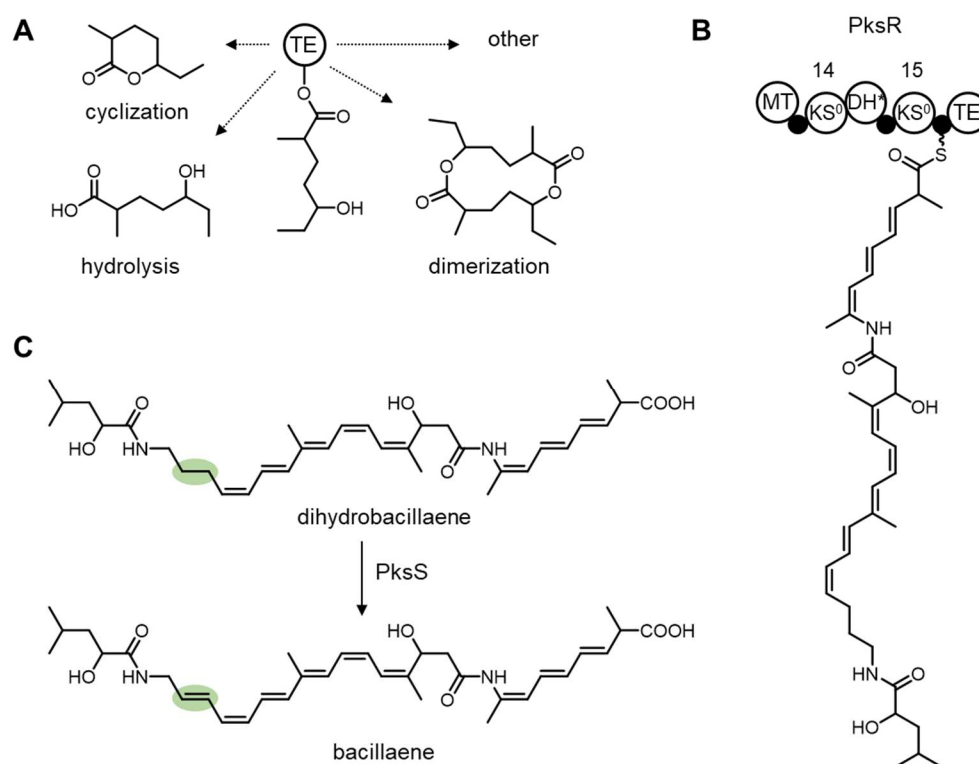
Figure 1. Versatility of thioesterase (TE) release mechanisms and the action of the terminus of the bacillaene *trans*-AT PKS. A) Offloading mechanisms commonly observed by TEs. B) Terminal domains of the *pksX trans*-AT PKS and the bound final intermediate. MT: methyltransferase, KS^0: non-elongating ketosynthase, DH*: dehydratase catalyzing double bond migration. Black circles indicate carrier proteins. C) After release by the TE, the cytochrome P450 enzyme PksS desaturates the precursor polyketide dihydrobacillaene (DHB) into the final product bacillaene. The point of desaturation is highlighted in green.

Results and Discussion

The *trans*-AT PKS cluster *pks*X of bacillaene was one of the first to be studied and the biosynthetic pathway was elucidated in 2007.[14] In *B. subtilis* DK1042, an assembly line consisting of two NRPS and 15 *trans*-AT PKS modules forms a bacillaene precursor by successive elongations of the assembly line-bound intermediates (Figure 1B, an overview of the bacillaene *trans*-AT PKS is provided in Figure S1). After offloading of dihydrobacillaene (DHB) by the PksR-TE, the polyketide is converted to the final natural product bacillaene by the discrete cytochrome P450 dehydrogenase PksS (Figure 1C).[15] To facilitate analysis of PKS functionality, we deleted the corresponding gene *pksS*, generating Δ*pksS*. This way, DHB can be measured as the sole end product of the cluster. The terminal modules used in this study originate from a range of gene clusters and species (see Table S7) and include various types of offloading reactions. These mechanisms include hydrolysis of linear products (nosperin[16]), aromatic ring formation (psymberin[17]) as well as dimerization (difficidin[18]). To compare a set of closely related PKSs that perform either dimerization or cyclization, we included the releasing domains of assembly lines responsible for production of different, structurally related actin-binding polyketides.[19] Luminaolide and misakinolide are dimers, whereas tolytoxin is the monomeric, macrocyclic version of misakinolide (Figure 2A).[19] In addition to TEs covalently attached to the PKSs, we utilized the free-standing C domain from the basiliskamide *trans*-AT PKS.[20] Engineered strains were generated by homologous recombination into the genome of *B. subtilis* DK1042, a strain that has a point mutation in *comI* resulting in increased natural competence, thus facilitating genetic modifications of the strain by homologous recombination.[21] For the different hybrid strains used in this study, the genomic region encoding the terminal ACP-TE of a donor cluster (Figure 2B) was fused downstream of the final ACP on the *pksR* gene of the bacillaene *trans*-AT PKS (Figure 2C). The upstream *pksR* sequence was used as one homology arm for targeted insertion of the region into the respective region in the genome of *B. subtilis* by homologous recombination and the region downstream of *pksS* as the second homology arm. In this way, the bacillaene *pksR*-TE as well as *pksS* are deleted and the PKS is connected via ACPs to the non-native release factors. For the difficidin hybrid with the terminal ACP-TE encoded on *difL* in the native host, this resulted in the mutant strain *B. subtilis difL*-ACP-ACP-TE, abbreviated as DifL-ACP2-TE. Analogously, *B. subtilis psyD*-ACP-ACP-TE (PsyD-ACP2-TE) was constructed for the psymberin hybrid, *B. subtilis misF*-ACP-ACP-TE (MisF-ACP2-TE) for the misakinolide hybrid, *B. subtilis ttoF*-ACP-PCP-TE (TtoF-ACP2-TE) for the tolytoxin hybrid, *B. subtilis lumE*-ACP-ACP-TE (LumE-ACP2-TE) for the luminaolide hybrid, and *B. subtilis nspD*-ACP-ACP-TE (Nsp-ACP2-TE) for the nosperin hybrid. For the free-standing C domain from the basiliskamide *trans*-AT PKS, the corresponding gene *basC* was integrated into the *amyE* locus on the genome of *B. subtilis* to create *B. subtilis basC* (BasC) (Figure 2D). As a negative control, we included *B. subtilis* Δ*pksR*-TE (ΔTE), wherein offloading of DHB is impaired due to the absence of the terminal module. *B. subtilis* Δ*pksS* (WT), wherein offloading occurs as in the native cluster but only DHB is produced, was used as an expression control. An overview of the clusters and domain architectures used for hybrid construction is provided in Tables S1 and S7.
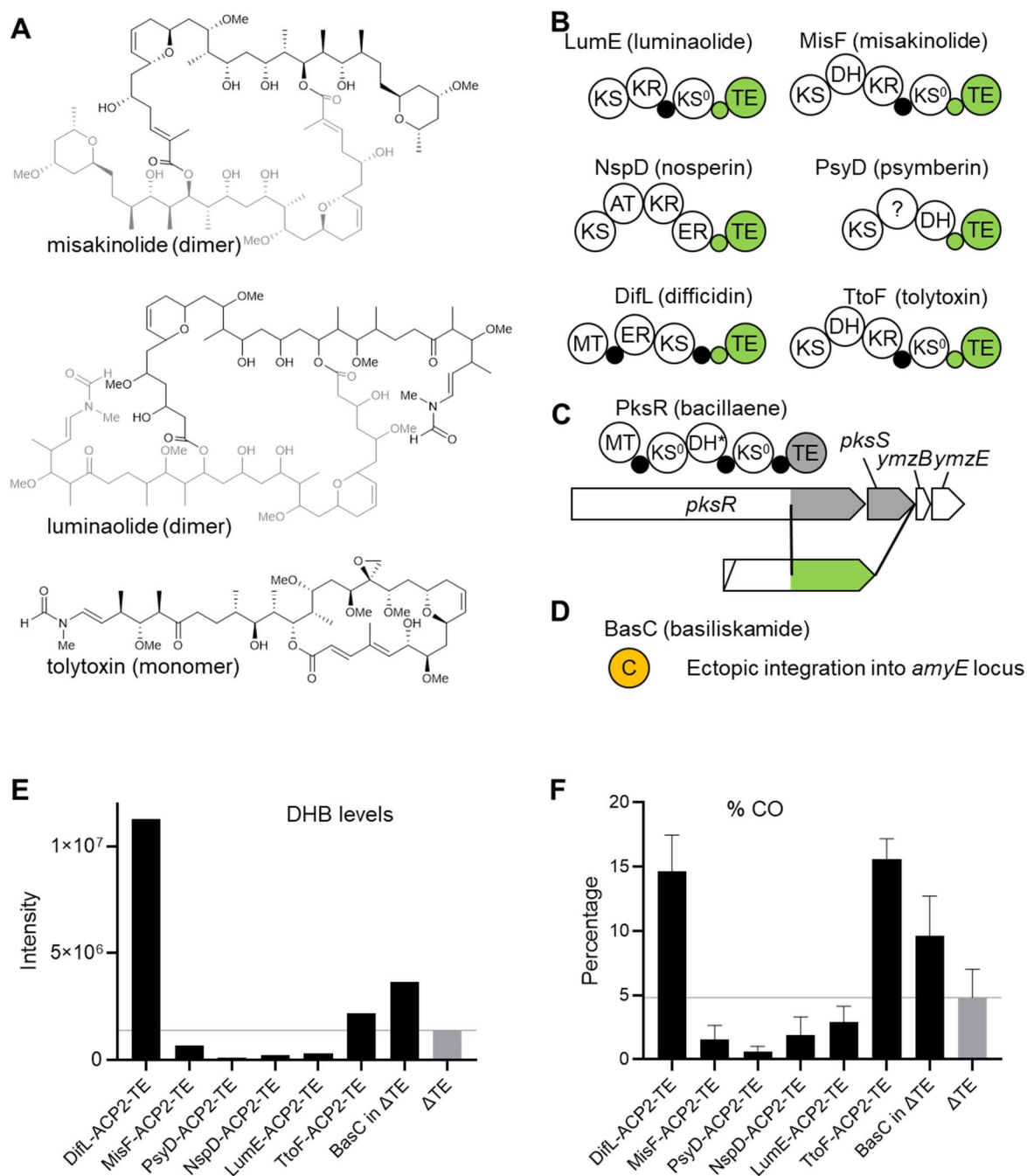
Figure 2. Hybrid cloning strategy and UHPLC-HRMS analysis of the TE swap mutants. A) Overview of the actin-binding polyketides used. Grey molecule parts represent a second monomer unit in dimers. B) Overview of the ACP-TE portions (marked in green) used in this study. Additional upstream (non-colorized) domains of each assembly line are shown as context. C) General principle to produce ACP-TE mutants. The TE-encoding region as well as the downstream *pksS* gene are exchanged for a non-native genetic region (indicated in green) through homologous recombination. D) The *trans*-acting BasC domain (marked in orange) was integrated into the genome of *B. subtilis* via ectopic integration into the *amyE* locus. E) DHB levels for the mutants that were coupled after the native PksR-ACP as well as the construct of ΔTE with *basC*. The ΔTE sample is plotted as a control. All DHB levels were normalized to bacillibactin levels. The horizontal line indicates the ΔTE levels of DHB. F) % total cluster output (% CO) and standard deviation computed for the mutants and ΔTE. The horizontal line indicates the % CO computed for ΔTE.

The hybrid strains were grown under conditions known to produce bacillaene, extracted with methanol, and subjected to ultra-high performance liquid chromatography/high-resolution mass spectrometry (UHPLC-HRMS). In addition, UV spectra were recorded to detect the production of bacillaene and intermediates (Figure S2). To account for variations in *Bacillus* biomass during cultivation, the production levels of the siderophore bacillibactin was assessed (Figure S3). For quantitative comparisons between the strains, DHB levels were normalized in reference to the wild type bacillibactin level as described in Jensen et al.[22] Functional bacillaene offloading was observed as DHB, which corresponds to the released final intermediate as shown in Figure 1B and C. UHPLC-HRMS analysis of extracts from hybrid strain cultures showed that most grafted modules demonstrated reduced bacillaene offloading compared to ΔTE (Figure 2E). Notable exceptions are the BasC, TtoF-ACP2-TE and DifL-ACP2-TE hybrid strains, wherein higher levels of offloaded DHB are found than in ΔTE. In addition to DHB level analysis, the amount of DHB relative to the total cluster output was calculated as percentage of overall cluster output (% CO) to quantify offloading efficiency (Figure 2F). We defined the total cluster output as the sum of all measured intermediates and DHB. With the expression control *B. subtilis* Δ*pksS* showing 53% CO, the unaided offloading from the assembly line in the ΔTE mutant was defined as a neutral result (4.8%). All other results were classified as either beneficial (aiding offloading, > 4.8% CO) or detrimental (impairing offloading, < 4.8% CO) (Table 1). The beneficial mutants are DifL-ACP2-TE (14.6%), TtoF-ACP2-TE (15.6%) and the BasC mutant (9.6%). This translates into 2- to 3.2-fold increased offloading over the ΔTE mutant but does not reach wild-type levels.

Table 1. Overview of DHB levels compared to ΔTE, % total cluster output (% CO) and fold increases. The number of replicates is given in parentheses. Mutants are classified as beneficial for increased and detrimental for decreased % CO compared to ΔTE. More detailed statistics are provided in Table S8.

Strain	% CO	Increase	Effect
DifL-ACP2-TE	14.6 ± 2.8% (4)	3.0	beneficial
BasC in ΔTE	9.6 ± 3.1% (7)	2.0	beneficial
MisF-ACP2-TE	1.5 ± 1.1% (4)	0.3	detrimental
PsyD-ACP2-TE	0.6 ± 0.4% (4)	0.1	detrimental
NspD-ACP2-TE	1.9 ± 1.4% (6)	0.4	detrimental
Lum-ACP2-TE	2.9 ± 1.2% (5)	0.6	detrimental
TtoF-ACP2-TE	15.6 ± 1.6% (3)	3.2	beneficial
Δ*pksS*	53.0 ± 13.7% (14)	11.0	+ control
ΔTE	4.8 ± 2.2% (15)	1.0	reference

To assess whether the choice of fusion site and the inclusion or exclusion of additional domains affects offloading, fusion sites ranging from the KS^0 of the penultimate bacillaene module to after the ACP directly upstream of the TEs were tested (Figure 3A and Figure S4). Such sets were generated for the active *pksR/ttoF* and *pksR/difL* hybrids in an effort to improve yields. For the seemingly detrimental hybrids *pksR/misF* and *pksR/psyD*, we generated the fusion site series to check whether offloading of DHB could be rescued by choosing a different site of recombination. The additional mutants were grown, extracted, and analyzed as described previously. While no increases in offloading activity were observed, UHPLC-HRMS analysis revealed a consistent pattern: In all four cases, the highest DHB levels were observed for the recombinations where the native bacillaene ACP was concatenated with the ACP native to the grafted TE (DifL-ACP2-TE, TtoF-ACP2-TE, MisF-ACP2-TE, PsyD-ACP2-TE) (Figure 3B and C). Detailed UHPLC-HRMS analysis is provided in Figures S5-11 and % CO computations are shown in Figure S12 and Table S8. Preservation of only one of the ACPs or coupling of the native bacillaene ACP to the tandem ACP in difficidin led to intermediate bacillaene levels. The lowest outputs were observed for mutants that contained an additional linker region (DifL-linker, MisF-linker) or a replaced KS^0 (MisF-KS^0). Intrigued by the activity of the free-standing BasC as a potential

universal offloading domain, we introduced BasC into WT *B. subtilis* to analyse whether WT titers could be increased. Levels of bacillaene production showed a clear increase, similar to our observations in the ΔTE mutant (Figure 3D). The % CO increased from 53.0% in the WT expression control to 76.6% translating to a 1.4-fold increase.
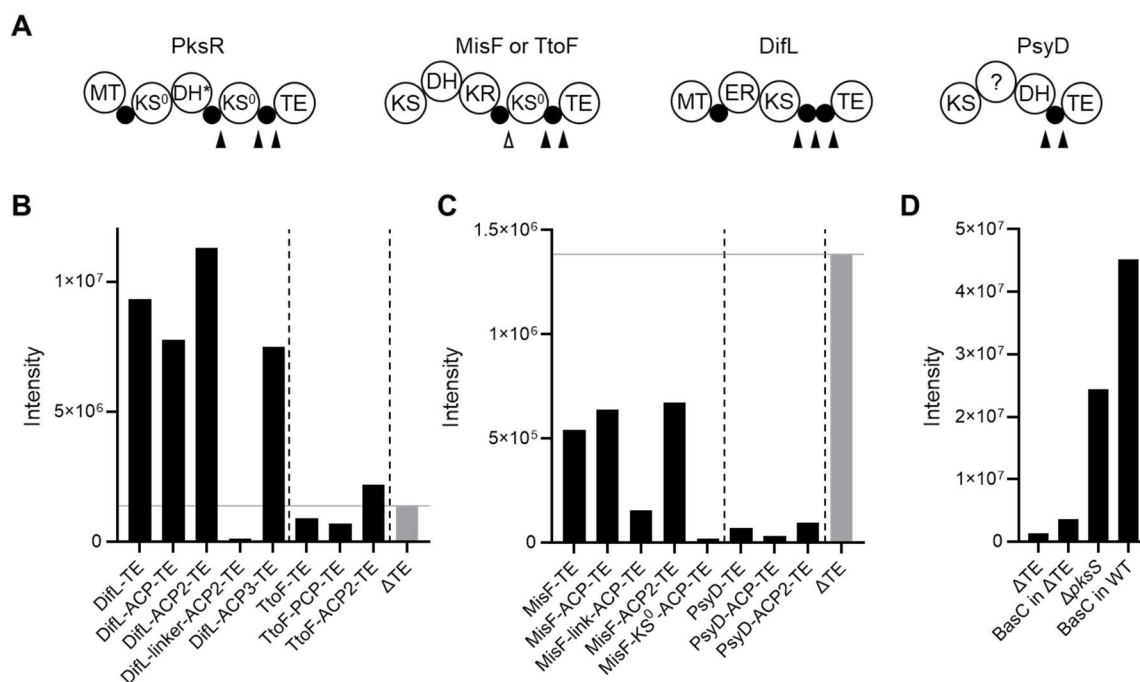

Figure 3. DHB levels normalized to bacillibactin levels for the different fusion site variants of the terminal hybrids. A) Overview of the TE swaps performed. Protein names are listed above the assembly lines. Black triangles indicate fusion sites. The empty triangle indicates the respective fusion was only used for MisF. A complete overview of the mutants is provided in Figure S4. B) Overview of the DHB levels for variants of the terminal hybrids that promote offloading. As a control, ΔTE is plotted. The horizontal line indicates the levels of bacillaene in ΔTE. C) Overview of the variants tested for the terminal hybrids. As a control, the production observed in ΔTE is plotted. The horizontal line indicates the levels of bacillaene in ΔTE. D) Results for the experiments of BasC in ΔTE and WT Δ*pksS*.

To assess transcription of the hybrid PKS, we performed mRNA extraction and subsequent cDNA synthesis of selected hybrid constructs of all functionality levels. For all tested constructs, bands referring to transcribed TE genes were observed (Figure S13). For hybrid strains that contained offloading modules with macrocyclizing or dimerizing activity in their natural context, we analyzed the extracts for potentially cyclized or dimerized bacillaene congeners. No such congeners were detected by UHPLC-HRMS analysis. As neither dimer formation nor cyclization are naturally expected for bacillaene but could theoretically happen based on the topology, we anticipated such reactions might occur at a slower rate. We aimed to prolong the reaction time through deletion of *pksB*, a gene encoding for an acylhydrolase (AH) in the bacillaene PKS. Through the deletion, we suspect that intermediates are kept on the TE for longer which could facilitate the occurrence of non-native reactions.[22-23] However, no dimerized or cyclized bacillaene congeners were observed upon bacillaene AH deletion.

Conclusion and Outlook

In this work, we wished to obtain first insights into the reprogramming potential of *trans*-AT PKSs. We assessed the exchangeability of extended terminal releasing modules in the biosynthetic pathway of bacillaene. The *pksR*-TE locus was exchanged by offloading domains from seven different *trans*-AT PKS pathways and culture extracts were analyzed by UHPLC-HRMS. Along with a putative *trans*-acting C domain from the basiliskamide *trans*-AT PKS, TEs with variable upstream domains were analyzed. These originate from different organisms and facilitate a range of release mechanisms. In their native context, these are hydrolysis of a linear polyketide, macrolactonization or aromatic cyclization, and dimerization. Using more closely related organisms, a tendency of higher success probability was observed. Difficidin and basiliskamide are naturally produced in *Bacillus* and *Brevibacillus* strains, respectively, and hybrids using parts from the corresponding *trans*-AT PKSs promote offloading. Except for the tolytoxin PKS-derived mutants, the other sets demonstrated generally lower PKS turnover. As successful PKS transcription was found via mRNA analysis, this could mean that additional inter-species differences could be responsible for these results. For example, codon usage or unfavorable mRNA secondary structures could alter translation.[24]

From a mechanistic perspective, it is interesting to see that TEs with a cyclizing mechanism have the ability to release linear bacillaene (Figure 4A). A non-hydrolytic release is traditionally thought to rely on an adjustment in the TE active site to shield the bound polyketide from water.[9] In this case, the TEs in our hybrid systems demonstrated the ability to release a linear product from an active site, suggesting that it is not completely shielded. A suboptimal interaction of the TE with the substrate could cause water to access the active site, thus promoting hydrolysis and release of linear DHB. This may explain why the cyclizing *difL*-TE readily offloads linear DHB. When comparing the set of actin-binding polyketides, it is interesting to see, that only the tolytoxin PKS-derived mutants promoted DHB release. TE grafts from the closely related misakinolide and luminaolide PKSs were much less active. In a phylogenetic tree, the TEs form a distinct clade within the *trans*-AT PKSs with the tolytoxin TE in the most basal position (Figures 4C and S14).[19] Considering that tolytoxin is a monomer while luminaolide and misakinolide are dimers, it is possible that the latter specialized their active sites alongside the new product during the course of evolution. According to the "waiting room" hypothesis, one monomer stays in the TE active site until the second arrives, prompting dimerization.[9] This waiting room would be present in luminaolide and misakinolide TEs, within which the substrate is attached. As DHB is unlikely to form a dimer, the assembly line would be stalled with the polyketide stuck in the TE active site. In contrast, the tolytoxin TE might have a more relaxed active site which could allow for release of DHB.

Next, it should be considered that the non-native TEs naturally accept substrates that look very different from the polyene DHB. It has been shown for some KSs of *trans*-AT PKS that they exhibit substrate selectivity based on the α- and β-positions of the incoming intermediate.[7, 25] The same positions on the TE bound intermediate are important for its stabilization during the release,[11] thus it can be hypothesized that their biochemical properties significantly contribute to the observed results. Structurally, TEs harbor an extended substrate channel that is flexible enough to accommodate substrates of different chain lengths.[10] In case of bacillaene, basiliskamide, tolytoxin, and misakinolide, the region close to the TE is composed of a reduced carbon chain, and in difficidin, a reduced chain with a β-exomethylene group is present. In psymberin, luminaolide, and nosperin, more polar keto or hydroxyl moieties are found. Considering these differences, it can be hypothesized that TEs accepting substrates similar to the native, non-polar bacillaene precursor should be preferred. In the case of TEs selective for polar, less-reduced carbon chains, the active site may contain polar residues. If confronted with a non-polar substrate, this could lead to suboptimal positioning of the intermediate and hampered

release (Figure 4B), as observed for the TE for linear substrate nosperin. Its polarity and bulky residues near the active-site serine might be disruptive to the TE's functionality.
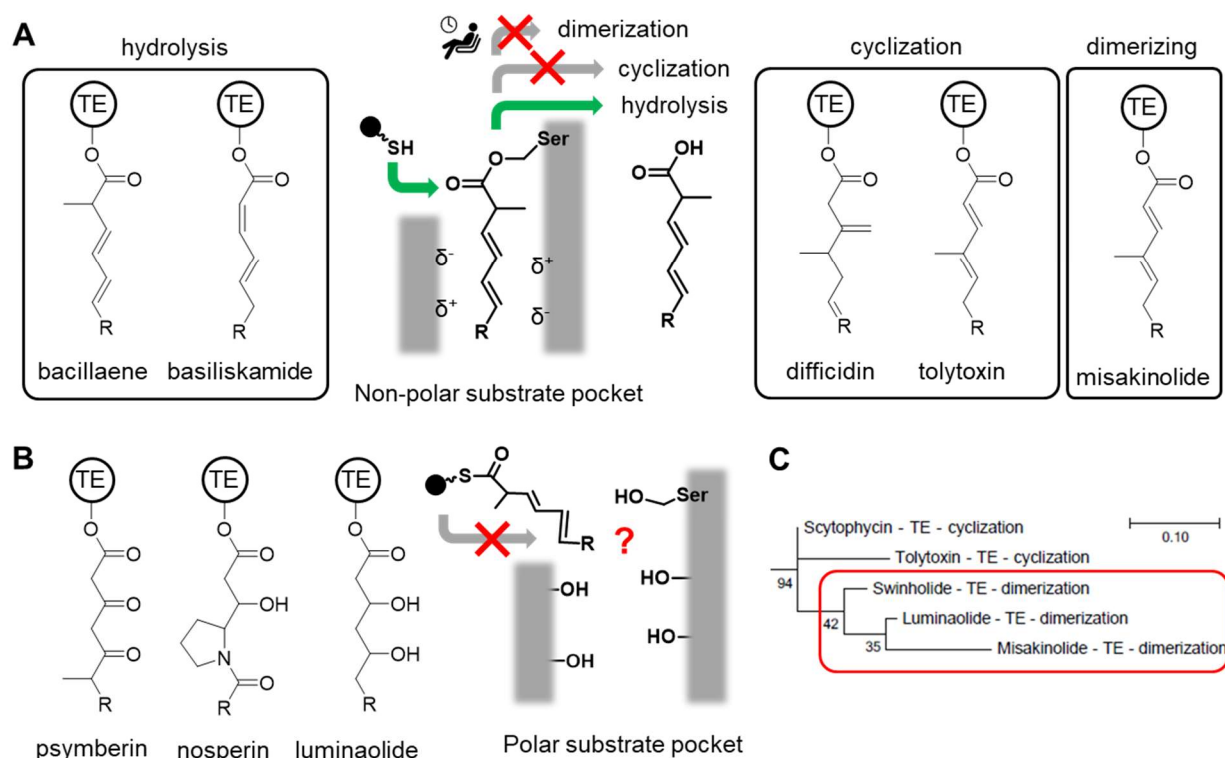

Figure 4. Factors affecting offloading by the TE. A) Overview of non-polar substrates accepted by TEs grouped by mode of release in their natural context. In these, the substrate channel contains non-polar residues which promote substrate transfer from the ACP (black circle) to the active site serine. The presence of water facilitates hydrolysis while cyclization and dimerization are impeded. B) TEs accepting polar final intermediates. The substrate channel cannot accommodate the non-polar bacillaene intermediate. C) Excerpt of a TE phylogenetic tree showing the closely related actin-binding polyketide producing PKSs and their mode of release. R: residue as in the polyketide, TE: thioesterase, wavy line: 4'-phosphopantheteinyl arm.

Another structural aspect is the detrimental effect of additional linkers. As TEs are expected to form dimers,[10] additional unstructured regions could impede dimerization and thus functionality. In general, our results indicate that a recombination site between one native and one non-native ACPs is favorable. While the exact positioning of ACPs during offloading is not known, it is possible that if concatenated, the substrate is shuttled from one ACP to the next prior to loading onto the TE.[26] Thereby the non-native proteins would only interact via ACPs, and native ACP-TE as well as ACP-upstream domain interactions would be conserved. [11] Recent studies have indicated that a substrate bound to the ACP is not freely accessible.[11] It is possible that in strains with a non-native TE, the substrate would be more buried, preventing AHs from accessing it. This could explain why some of the detrimental mutants show DHB levels that are even lower than in a TE deletion mutant. In the latter, the lack of a TE may make the ACP more accessible due to impeded dimerization and thus PKS quaternary structure. In the case of the triple ACP mutant DifL-ACP3-TE the increased distance may be the cause of a reduced activity compared to DifL-ACP2-TE.

Lastly, the results obtained from free-standing BasC indicate that this condensation domain releases final intermediates from a non-cognate *trans*-AT PKS assembly line *in trans*. As mentioned above, the native *Brevibacillus* system may be advantageous on the level of protein expression. Still, the observed flexibility renders the discrete enzyme a highly attractive candidate as a universal tool in *trans*-AT PKS cluster engineering to overcome the TE bottleneck.[13] Overall our results demonstrate that the task of finding the right TE for an engineered cluster is a multi-layered challenge. This work underlines the necessity to preserve the ACP native to the TE and the importance of removing additional sequences. The extra linker regions were thought to allow extra flexibility or act as an adapter (in the case of KS^0) but were ultimately detrimental to the function of the TE hybrids. Furthermore, we provide evidence that mixed ACP combinations act as adapters and can even improve the output of barely functional hybrid clusters. Considering the exceptional offloading activity of BasC, this protein could be used in PKS hybrids impaired by an incompatibility of product and TE domain. All in all, we provide the first examples and strategy to tackle the engineering of *trans*-AT PKS clusters which ultimately rely on the capability of the TE to offload a non-native substrate.

Acknowledgements

References

[1] a) D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2020**, *83*, 770-803; b) M. S. Butler, A. A. B. Robertson, M. A. Cooper, *Nat. Prod. Rep.* **2014**, *31*, 1612-1661; c) K. J. Weissman, *Nat. Prod. Rep.* **2016**, *33*, 203-230.

[2] C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688-4716.

[3] a) E. J. Helfrich, J. Piel, *Nat. Prod. Rep.* **2016**, *33*, 231-316; b) J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996-1047; c) M. L. Adrover-Castellano, J. J. Schmidt, D. H. Sherman, *ChemCatChem* **2021**, *13*, 2095-2116.

[4] P. F. Leadlay, *Curr. Opin. Chem. Biol.* **1997**, *1*, 162-168.

[5] A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, *Chem. Rev.* **2019**, *119*, 12524-12547.

[6] E. J. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, *Nat. Commun.* **2021**, *12*, 1-14.

[7] T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225-233.

[8] A. A. Koch, J. J. Schmidt, A. N. Lowell, D. A. Hansen, K. M. Coburn, J. A. Chemler, D. H. Sherman, *Angew. Chem. Int. Ed.* **2020**, *59*, 13575-13580.

[9] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255-278.

[10] S. Smith, S.-C. Tsai, *Nat. Prod. Rep.* **2007**, *24*, 1041-1072.

[11] M. Wang, C. N. Boddy, *Biochemistry* **2008**, *47*, 11793-11803.

[12] a) S.-C. Tsai, H. Lu, D. E. Cane, C. Khosla, R. M. Stroud, *Biochemistry* **2002**, *41*, 12598-12606; b) M. E. Horsman, T. P. A. Hari, C. N. Boddy, *Nat. Prod. Rep.* **2016**, *33*, 183-202.

[13] D. A. Hansen, A. A. Koch, D. H. Sherman, *J. Am. Chem. Soc.* **2017**, *139*, 13450-13455.

[14] a) J. Moldenhauer, X. H. Chen, R. Borriss, J. Piel, *Angew. Chem. Int. Ed.* **2007**, *46*, 8195-8197; b) R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straightt, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506-1509.

[15] J. J. Reddick, S. A. Antolak, G. M. Raner, *Biochem. Biophys. Res. Commun.* **2007**, *358*, 363-367.

[16] A. Kampa, A. N. Gagunashvili, T. A. Gulder, B. I. Morinaka, C. Daolio, M. Godejohann, V. P. Miao, J. Piel, O. Andresson, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3129-3137.

[17] K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews, J. Piel, *Nat. Chem. Biol.* **2009**, *5*, 494-501.

[18] X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, G. Gottschalk, R. D. Süssmuth, R. Borriss, *J. Bacteriol.* **2006**, *188*, 4024-4036.

[19] R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. N. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, S. Matsunaga, J. Piel, *Nat. Chem. Biol.* **2015**, *11*, 705-712.

[20] C. M. Theodore, B. W. Stamps, J. B. King, L. S. Price, D. R. Powell, B. S. Stevenson, R. H. Cichewicz, *PLoS One* **2014**, *9*, e90124.

[21] M. A. Konkol, K. M. Blair, D. B. Kearns, *J. Bacteriol.* **2013**, *195*, 4085-4093.

[22] K. Jensen, H. Niederkruger, K. Zimmermann, A. L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Poplau, C. Kohlhaas, C. A. Townsend, M. Oldiges, C. Hertweck, J. Piel, *Chem. Biol.* **2012**, *19*, 329-339.

[23] S. Kosol, M. Jenner, J. R. Lewandowski, G. L. Challis, *Nat. Prod. Rep.* **2018**, *35*, 1097-1109.

[24] C. Gustafsson, S. Govindarajan, J. Minshull, *Trends in Biotechnol.* **2004**, *22*, 346-353.

[25]   E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, *Nat. Chem. Biol.* **2019**, *15*, 813-821.

[26]   A. Ranganathan, M. Timoney, M. Bycroft, J. Cortés, I. P. Thomas, B. Wilkinson, L. Kellenberger, U. Hanefeld, I. S. Galloway, J. Staunton, *Chem. Biol.* **1999**, *6*, 731-741.

Experimental procedures

General

All restriction endonucleases used were obtained from New England BioLabs (Ipswich, MA, USA). For Gibson assembly, the Gibson assembly kit from New England BioLabs (Ipswich, MA, USA) was used. T4 ligase and buffer for Golden Gate (GG) cloning were obtained from Promega (Fitchburg, WI, USA). Colony PCR was performed with Phusion polymerase (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. DNA sequencing was conducted by GATC Biotech (Konstanz, Germany) or Microsynth (Balgach, Switzerland). The pJET1.2/blunt and pCR-Blunt II-TOPO cloning kits were purchased from Thermo Fisher Scientific (Waltham, MA, USA) and the pGEM-T Easy subcloning kit was obtained from Promega (Fitchburg, WI, USA).

Strains and culture conditions

We used *B. subtilis* DK1042 which is a derivative of the *B. subtilis* 3610 wild-type strain with increased competence due to a single point mutation in the *comI* gene negatively regulating natural competence.[1] An overview of all strains used in this work is provided in Table S1. Strains were grown in LB liquid medium and on LB agar plates at 37 °C. To select plasmids in *E. coli*, antibiotics were used at final concentrations as follows: ampicillin (100 µg/mL), chloramphenicol (12.5 µg/mL), spectinomycin (25 µg/mL), and gentamycin (20 µg/mL). To select for positive transformants in *B. subtilis*, antibiotics were used at final concentrations as follows: spectinomycin (60 µg/mL or 100 µg/mL) and chloramphenicol (5 µg/mL). To test for polyketide production, *Bacillus* strains were grown in YEME7 medium[2] at 37 °C in darkness.

Plasmid construction

As an acceptor vector, pFusA was used. The plasmid was a gift from Adam Bogdanove & Daniel Voytas (Addgene plasmid #31028). All gene fragments and resistance genes used in Golden Gate (GG) cloning were amplified with primers containing the *Bsa*I restriction sites (Table S2). All PCR fragments were subcloned into pJET1.2/blunt, pCRBlunt II-TOPO or pGEM-T Easy according to manufacturer's instructions. GG was performed as published by C. Engler and S. Marillonnet using pFusA as an acceptor vector.[3] See Table S3 for a list of plasmids and information on donor and acceptor vectors for GG assembly. The chloramphenicol (Cm) resistance gene was amplified from pACYC184 (New England BioLabs, Ipswich, MA, USA) and subcloned into pCR-Blunt II-TOPO. The HyperSpac promoter was amplified from pMF37/pDGICZ and fused to the Cm resistance gene using Gibson assembly. Subcloned GG pieces and GG constructs were introduced into *E. coli* DH5α. As a common acceptor plasmid, pFusA with a placeholder kanamycin resistance cassette, the chloramphenicol resistance cassette and a common downstream homology was assembled. For assembly of the hybrid constructs, codon-optimized synthetic DNA was acquired from GenScript (sequences are provided in Table S4). The synthetic DNA was inserted into the common acceptor plasmid pFusA/S/C_KanR by restriction cloning with PmeI, NheI and SpeI (Table S5). The final assemblies were introduced into *E. coli* DH5α.

Bacillus transformation and screening

The transformation protocol was obtained from http://2012.igem.org. 3 mL of freshly prepared medium A (25% glucose 1 mL, 10x medium A (yeast extract 5 g, casamino acids 1 g, ddH$_2$O add 450 mL) 4 mL, 10x salt solution (NH$_4$)$_2$SO$_4$ 10 g, K$_2$HPO$_4$ 69.8 g, KH$_2$PO$_4$ 30 g, Na$^+$ citrate 5 g, MgSO$_4$·7 H$_2$O 1 g, ddH$_2$O add 500 mL) 4.5 mL, H$_2$O ad 45 mL) were inoculated with a single colony from a fresh plate and incubated at 37 °C and 180 rpm overnight. 20 mL of freshly prepared and pre-warmed (37 °C) medium A was inoculated to an OD$_{600}$ of 0.1 with the overnight culture. Cells were grown to an OD$_{600}$

of 1.0 at 37 °C and 180 rpm. 45 mL of prewarmed freshly prepared medium B (25% glucose 1 mL, 10x medium A 4 mL, 10x salt solution 4.5 mL, H_2O ad 45 mL, 1 M $CaCl_2$ 22.5 µl, 1 µM $MgCl_2$ 112.5 µl, ddH$_2$O add 45 mL) were inoculated with 5 mL of the culture and incubated for 1.5 h at 37 °C and 180 rpm. Cells were harvested by centrifugation for 5 min at 1800 × g and at RT and resuspended in 3 mL of supernatant. 1 µg of DNA was added to 100 µl of cell suspension. After 2 h of incubation at 37 °C and 900 rpm cells were plated on LB containing 5 mg/mL of Cm or 60 mg/mL Spc, depending on the transformed construct. Plates were incubated overnight at 37 °C. Single colonies were picked and transferred to a fresh LB plate containing appropriate antibiotics and incubated overnight at 37 °C. If a mutant already containing one of the selection markers was transformed, colonies were plated on both antibiotics separately for counter selection screening. Positive colonies were chosen for colony PCR. *Bacillus* colony material was cooked for 10 min in 20 µl of DMSO and used as a template. An overview of screening primers is provided in Table S6.

mRNA extraction
mRNA extraction and cDNA synthesis were performed using the RNeasy Mini Kit and the Omniscript RT Kit from Qiagen according to the manufacturer's instructions. Presence of TE transcripts was assessed by colony PCR. Primers used in colony PCR are provided in Table S6.

Sequence alignments of TE and offloading C sequences
Amino acid sequences of 78 TEs or offloading C domains from an in-house database of 78 annotated *trans*-AT PKS clusters. As an outgroup, the TE from the erythromycin *cis*-AT PKS was used. The sequences were aligned using the MUSCLE algorithm with default settings[4] and a phylogenetic tree was computed with MEGA7[5].

Expression and analysis of bacillus hybrids
For expression, 5 mL YEME7 medium (4 g/L yeast extract, 10 g/L malt extract, 4 g/L glucose, 0.5 mL of 1 M NaOH per liter, add MOPS after autoclaving) with the appropriate antibiotics was inoculated with a single colony of *B. subtilis* strains. The culture was incubated in the dark at 37 °C and 250 rpm for 16-18 h. To check for dehydrobacillaene (DHB) production or intermediates, the culture was mixed in the dark with methanol and shaken vigorously. The samples were spun down for 20 min at 10 °C and max speed and the supernatant subjected to UHPLC-HRMS analysis. Kinetex 2.6 µm XB-C18 100 A column (4.6 × 150 mm; Phenomenex) for 5 min at 10% acetonitrile (0.1% FA) in water (0.1% FA) followed by a gradient (10 to 100%) of acetonitrile (0.1% FA) in water (0.1% FA) over 11.5 min followed by 2 min 10% acetonitrile (0.1% FA) in water (0.1% FA) (1 mL/min). HPLC-ESI/MS spectra were obtained from a Q-Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer set to positive mode and coupled to a Dionex UltiMate™ 3000 UHPLC System (Thermo Scientific).

Supplementary Tables and Figures

Table S1. Strains used and created in study.

Strain	Characteristics or sequence (5' → 3')
E. coli DH5α (Invitrogen)	F⁻ Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR*17 (rK⁻, mK⁺) *pho*A *supE44* λ⁻ *thi*-1 *gyrA96 relA1*
B. subtilis DK1042 [6]	*B. subtilis* 3610 *comI*^{Q12L}
*B. subtilis pks*TE_KO	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *Cm^R, comI*^{Q12L}
*B. subtilis pksS*_KO	Δ⁻ *pksS, ymzB*, Δ⁺ *Cm^R, comI*^{Q12L}
*B. subtilis pks*_KO	Δ⁻ *pks*, Δ⁺ *Cm^R, comI*^{Q12L}
B. subtilis BasC	Δ⁺ *comI*^{Q12L}, *amyE:: basC, Spc^R*
*B. subtilis DifL*_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *DifL*_TE, *Cm^R, comI*^{Q12L}
*B. subtilis DifL*_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *DifL*_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis DifL*_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *DifL*_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis DifL*_link_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *DifL*_link_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis DifL*_ACP_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *DifL*_ACP_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis MisF*_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *MisF*_TE, *Cm^R, comI*^{Q12L}
*B. subtilis MisF*_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *MisF*_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis MisF*_link_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *MisF*_link_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis MisF*_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *MisF*_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis MisF*_KS⁰_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *MisF*_KS$_0$_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis PsyD*_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *PsyD*_TE, *Cm^R, comI*^{Q12L}
*B. subtilis PsyD*_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *PsyD*_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis PsyD*_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *PsyD*_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis NspD*_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *Nos*_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis LumE*_ACP_ACP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *LumE*_ACP_ACP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis TtoF*_PCP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *TtoF*_PCP_TE, *Cm^R, comI*^{Q12L}
*B. subtilis TtoF*_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *TtoF*_TE, *Cm^R, comI*^{Q12L}
*B. subtilis TtoF*_ACP_PCP_TE	Δ⁻ *pksR*_TE, *pksS, ymzB*, Δ⁺ *TtoF*_ACP_ACP_TE, *Cm^R, comI*^{Q12L}

Table S2. Primers used in this study.

Name	Sequence	Template	(sub)cloned
Gent DraIII Fw	ATCACACCGTGTTAGGTGGCGGTAC	TREX vector pIC20H-RL JX668229.1[7]	pFus_A/*Spc*
Gent DraIII Rv	ATC ACGTA GTGTAGGGATAACAGGGTAA		
GG SpeR2 Fw	TTACTAGTTAACCATCGTGACGCGGC	pIC333[8]	pGEM_GG_*Spc*R
GG SpeR2 Rv	ATCCTGCAGGCTAATTGAGAGAAGTTTC		
CmR2.2Fus Fw	CTTACATAAGGAGGAACTACTATGGAGAAAAAAATCACTGGA	pACYC184 New England BioLabs (Ipswich, MA, USA)	pTopo_Cm^R
GG CmR2.2 Rv	TGGTCTCATACGTTACGCCCCGCCC		
Phspac Fw	AGTTTAAACTACACAGCCCAGTCCAGACT	pMF37/pDGICZ EU864235.1[9]	pTopo_GG_pHCmR
pHspacFus Rv	TCCAGTGATTTTTTTCTCCATAGTAGTTCCTCCTTATGTAAG		
GG pHspac Fw	AGGTCTCAACTCGTTTAAACTACACAGCCCAGT	pTopo_CmR	
GG CmR2.2 Rv	TGGTCTCATACGTTACGCCCCGCCC		
GG pH CmR3 Fw	AGGTCTCAAGCTGTTTAAACTACACAGCCCAG	pTopo_GG_pHCmR	pTOPO_pHCmR3
GG CmR2.2 Rv	TGGTCTCATACGTTACGCCCCGCCC		
KanR7 Fw	AGGTCTCACTAT ACTAGT CTGCGCTAGCATG	pCOLADuet-1 Novagen (Merck; Darmstadt, Germany)	pGEM_Kan^R
KanR7 Rv	AGGTCTCA GAGT TTAGAAAAACTCATCGAGCATCAAATG		
baeKO Bs Fw	AGGTCTCACTATCTTATGTCACGATTT	*B. subtilis* 3610 gDNA, DSMZ	pGEM_GG_pksKO
baeKO Bs Rv	TGG TCT CTA GCTCTGCAAAATTTATATA		
GG ymzB Fw	AGGTCTCACGTATGCACGATCTGTTACGA	*B. subtilis* 3610 gDNA, DSMZ	pJET_GG_ymzB
GG ymzBE Rv	TGGTCTCTCGCCTTAATTAACATCAAAACTGAACC		
Gi_TtoF-PCP-TE-F	CATTTGAAAAGAGAAGCTGCTGGGAGGCTAAAGTTGAGGTTCAAG	pFusA_*TtoF*_ACP_PCP_TE	pFusA_*TtoF*_PCP_TE
Gi_TtoF-PCP-TE-R	CTTGAACCTCAACTTTAGCCTCCCAGCAGCTTCTCTTTTCAAATG		
Gi_TtoF-TE-F	CATGATGGACTTGATCGCAAAGCAAAAGATACGCCGCAAATTG	pFusA_*TtoF*_ACP_PCP_TE	pFusA_*TtoF*_TE
Gi_TtoF-TE-R	CAATTTGCGGCGTATCTTTTTGCTTTGCGATCAAGTCCATCATG		
Gi_TtoF-vector-F	CTCTGGTTCAGTTTTGATGGGC	pFusA_*TtoF*_ACP_PCP_TE Gibson piece of the backbone for pFusA_*TtoF*_PCP_TE and pFusA_*TtoF*_TE	
Gi_TtoF-vector-R	GCCCATCAAAACTGAACCAGAG		

Table S3. Golden Gate assemblies performed in this study.

New construct	Backbone	Insert(s)	
pFusA/S/C_KanR	pFus_A/Spc	pGEM_GG_Kan^R pTopo_GG_phCm^R pJET_GG_ymzB	Spacer SM DGH
pFusA/G/S_KanR	pFus_A/Gm	pGEM_GG_Kan^R pGEM_GG_Spc^R pJET_GG_ymzB	Spacer SM DGH
pFusA/S_*pks*_KO	pFus_A/Spc	pGEM_GG_pksKO pTopo_GG_phCm^R pJET_GG_ymzB	VGH SM DGH
pFusA/S_*pks*_TEKO	pFus_A/Spc	pTOPO_GG_TE_KO pTopo_GG_phCm^R pJET_GG_ymzB	VGH SM DGH
pFusA_*psyD*_TE	pFus_A/Spc	pTOPO_*pksR*_ACP pTOPO_*psyD*_ACP_TE pTopo_GG_phCm^R pJET_GG_ymzB	VGH FGF SM DGH
pFusA_*psyD*_ACP_TE	pFus_A/Spc	pTOPO_*pksR*_nahvilee pTOPO_*psyD*_nahvilee_ACP_TE pTopo_GG_phCm^R pJET_GG_ymzB	VGH FGF SM DGH
pFusA_*psyD*_ACP_ACP_TE	pFus_A/Spc	pTOPO_*pksR*_ACP pTOPO_*psyD*_ACP_ACP_TE pTopo_GG_phCm^R pJET_GG_ymzB	VGH FGF SM DGH

Table S4. Sequences used in this work.

Insert(s)	
pUC57_*basC*	AAGCTTACATAAGGAGGAACTACTTTGGTGAAAAAAACGGTGATTAAAACGTATCCGTTATCAAAAGTGCAGAGTGTATTTTATCAACATTATGCCCTGCTTCCTGATTCCTATGGAAACAATGAAAAAATCCTTTACCGTATACATTCCAAGATCAATATAGATGTCATGAAGGAAGCCTTCTTACAGGTGATCAAGCGACATGAAATGTACCGTACCAATTTTCTTATGGAGGATGAGCCGGTGCATAAAGTATATGAGGAAGCGATTCTGGATTTTGAAGTGGTGGATGCTAGCGGTTGGTCACAAGAGCTCGTAACACAGTTTTTAAGTGAGGAAACCTACCGCCCCTTTCAATTGGAAGAAAATCCTCCCTTTCGAATTAGATTGTTCTGCTTTTCAGAGGATGATTTTATCTTGCTTTTAGTGTGGCATCATGTTGCTGTAGATGGCTGGTCTGCTTCTCTTACAATCGATGATTTTGGATTATTATACAAGAGTTTATTGGAAGGAAAGGAATGTGAGCTTACTCCGATCAGAAAACAGTTTTCAGACTTTGTTGTAGCCCAAAACGAAATGTTGGAAAGTCCAGAAGGTGAAAGACAGCGTTTATTTTGGAAAGAAAAGCTGTCAGGGGGAGCTTCCTGTCCTGAAGCTGCCAGCAGATCGAGATTTTCCACAAGTTCGTACGTATAAAGGAGGGACGGTATCGTTTCGTCTTGAGAAAACAGTAAGCGATCAAGTTTTTGCTTATTGTCAACAGAATGAGACGACCCTTGATCGAGTTCTGCTCTCCTTATATTTCGCGTTTTTACATGGTTATTCACAGCAAGAAGAAATCGTAATTGGAACACCTAGGTATGGACGTGCATCGACAAGCTATTATAATACGTGTGGTGTTTTTGTCAATCTAATTCCATTGCGATTGAAAATGCCGAAAAATGCCACGTATGCTGAGCTGGTGCATTTTGTTGATGAACAAATTAAACAATGTCTGGATTATCAGGATTTTCCTATGACGCTTATGATGGAAGAGCAGGATATCAAAACAGGTTTAACTCCCTTGTTTCAAACCTGCTTCTCCATTCAAAAATGGCCTAGACAAAACACAACGTTTCATTTTGGGGATTCCACACACGAGATCAATTTACATGGATTACAAATGAGCCACTACTATCAGAAAAGAAAATCTCAAAGTTTGATCTTGCGCTGTATGTGGAAGAAGATCGTAAGAAAGGCATTTTACCTATCTTTGAATATAATGCTGATATTTTTGATAAAAAAACGATAGAGAAGCTAAGTCGTATTTTCATAACATTCATCACTTCGGTTATGGAAGATGATCAGGTGAAAATTTCCACCTTGTGCTCATACCTTCAAGCGGTTGAGAAATAGGGATCC
pUC57_*difL*_TE	AATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACCGGCCGAGAGGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAAAGAGAAGCTGCTGGGAGTATGCTGAAGCGAGGAAAACAGCTGTACGTCCGCAAATCCCGGAAGCAGAAACTCCCCTTGCCGGCATCTCTGACGGCACTTTGCGCAAAGAGCTCAAAGATAGCCTTGCGGACATCCTGTTTCTGAAGCCGGAAGATATTGATGAGCATGAGGCATTCATTGAGATGGGGCTGGATTCTATTATCGGTGTGGAATGGGTTCAGTCCATTAATAAAACCTATCAAGCGTCTATTACGGCAAATCTGGTCATGAATATCCGACAATCGCAACATTAGCCGGGTATCTGACCGGTTCTGTCCAAAAAGACACCGAGGAAGTACATGGGGAACATACGAAACAAAACGATATATCAGAAGAGACAGATGATCAAGAACCGCCCGAGCTGATCTGTCTTAACCCGGGTGGAAAGGGAAGACCGATTTTCTGGTTTCATTCCGCGCTGGGGGGCAGTGGAGCTGTACCAGCCTGTTGCCGAACGGACCGGGCGTCCTTTTTACGGAATTCAGGCAAGAGGGTGGATGACAGACCGGGAACCGATCCGCGGCATTAAAAACATGGCCGAATATTACGTTAAGCTCATACGTACCGTGCAGCCCCGGGGTCCTTACGATGTAGGCGGCTATTCATTAGGGGGCATGCTCGCTTATGAAGTCACCCGGCAATTGCAGCTGGCTGGAGAAACCGTTGAGTCAATCGTCATGCTCGACTCCTTTGACTCGGCGGGTGCGCGGGAAATAAACATTCCTGATGAATTATTTGATAAAAATATGATGCTGCAGACGGTGAACATGGCTTTATTTGCACCAATCTATCAACATCCGGAACGCATTTCTGAAACCTTGATTCATCGTGATGAACTGAATGCCGATCTTCATGAAGACATATTTTTAGAGCAATTAATCGCAAAAGCCAAACAGCGGGGGTTAACGAAGACGGACGCGTTGGTGAAAAATCTGATCAGGCAGAATATAAACGTTCAGGACGCTTACGAAATCACTGATTACGAACTCCTGCCGCTTCCGGATCCGAACGGGGTGACATGCTATTATTTCCGAAATAAAAACGGTATTTTCTTGGGTGAGCTTGATCCGTATTTTACGATCAGGGCTGATATTTTTATCCGCTGATTTCGTCAACTATTGGAGCCAGTGGGAAGAGCATATGACAAACATCGCCATCATGGATGTGGCGGTATCCAACCATATGACGTTTATTGGCAGAAGACAGCACAGCCGAAACGATTTCTGAATTTTGCGGCGTTTTATATTCGGCAGACGGAATGACAGAATACAGCTTTCTGCAGCTGCAGGAAAAGATACTGAATGTGCATGGTGAGATGCTGAAGCAGCCCGCCAAAACAAGACCGTAA
pUC57_*difL*_ACP_TE	CCAAATCCGGCCGGCATGAAGGATGCAATGCTGAAGGCTTATCAAGGAGCGCAAATTGATCCCAAAACGGTGACCTATATAGAAGCGCATGGGATCGCCTCTCCATTGGCAGACGCGATAGAAATAGAGGCGTTAAAGTCAGGCTGCAGTCAGCTCGAATTGGAACTTCCACAGGAAGTACGGGAGGAAGCGCCATGTTATATCAGCAGCTTAAAGCCGAGCATCGGACACGGTGAACTCGTCTCAGGCATGGCTGCTCTTATGAAGGTCAGCATGGCGATGAAGCATCAAACAATACCAGGCATATCCGGATTTTCGTCTTTGAATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACCGGCCGAGAGGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAAAGAGAAGCTGCTGGGAGTATGCTGAAGCGAGGAAAACAGCTGTACGTCCGCAAATCCCGGAAGCAGAAACTCCCCTTGCCGGCATCTCTGACGGCACTTTGCGCAAAGAGCTCAAAGATAGCCTTGCGGACATCCTGTTTCTGAAGCCGGAAGATATTGATGAGCATGAGGCATTCATTGAGATGGGGCTGGATTCTATTATCGGTGTGGAATGGGTTCAGTCCATTAATAAAACCTATCAAGCGTCTATTACGGCAAATCTGGTCATGAATATCCGACAATCGCAACATTAGCCGGGTATCTGACCGGTTCTGTCCAAAAAGACACCGAGGAAGTACATGGGGAACATACGAAACAAAACGATATATCAGAAGAGACAGATGATCAAGAACCGCCCGAGCTGATCTGTCTTAACCCGGGTGGAAAGGGAAGACCGATTTTCTGGTTTCATTCCGCGCTGGGGGGCAGTGGAGCTGTACCAGCCTGTTGCCGAACGGACCGGGCGTCCTTTTTACGGAATTCAGGCAAGAGGGTGGATGACAGACCGGGAACCGATCCGCGGCATTAAAAACATGGCCGAATATTACGTTAAGCTCATACGTACCGTGCAGCCCCGGGGTCCTTACGATGTAGGCGGCTATTCATTAGGGGGCATGCTCGCTTATGAAGTCACCCGGCAATTGCAGCTGGCTGGAGAAACCGTTGAGTCAATCGTCATGCTCGACTCCTTTGACTCGGCGGGTGCGCGGGAAATAAACATTCCTGATGAATTATTTGATAAAAATATGATGCTGCAGACGGTGAACATGGCTTTATTTGCACCAATCTATCAACATCCGGAACGCATTTCTGAAACCTTGATTCATCGTGATGAACTGAATGCCGATCTTCATGAAGACATATTTTTAGAGCAATTAATCGCAAAAGCCAAACAGCGGGGGTTAACGAAGACGGACGCGTTGGTGAAAAATCTGATCAGGCAGAATATAAACGTTCAGGACGCTTACGAAATCACTGATTACGAACTCCTGCCGCTTCCGGATCCGAACGGGGTGACATGCTATTATTTCCGAAATAAAAACGGTATTTTCTTGGGTGAGCTTGATCCGTATTTTACGATCAGGGCTGATATTTTTATCCGCTGATTTCGTCAACTATTGGAGCCAGTGGGAAGAGCATATGACAAACATCGCCATCATGGATGTGGCGGTATCC

	AACCATATGACGTTATTGGCAGAAGACAGCACAGCCGAAACGATTTCTGAATTTTGCGGCGTTTTATATTCGGCAGACGGAATGACAGAATACAGCTTTCTGCAGCTGCAGGAAAAGATACTGAATGTGCATGGTGAGATGCTGAAGCAGCCCGCCAAAACAAGACCGTAA
pUC57_*difL*_ACP_ACP_TE	CCAAATCCGGCCGGCATGAAGGATGCAATGCTGAAGGCTTATCAAGGAGCGCAAATTGATCCCAAAACGGTGACCTATATAGAAGCGCATGGGATCGCCTCTCCATTGGCAGACGCGATAGAAATAGAGGCGTTAAAGTCAGGCTGCAGTCAGCTCGAATTGGAACTTCCACAGGAAGTACGGGAGGAAGCGCCATGTTATATCAGCAGCTTAAAGCCGAGCATCGGACACGGTGAACTCGTCTCAGGCATGGCTGCTCTTATGAAGGTCAGCATGGCGATGAAGCATCAAACAATACCAGGCATATCCGGATTTTCGTCTTTGAATGACCAGGTGTCATTAAAGGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACCGGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAAAGAGAAGCTGCTGGGCGGCTGATGTAAACAGTTCTTTACCAGGCCAAGCCATGAAAAAGTCTGAAACGCAAACCGTTCTCCCGTCTTCGCCGCAGCCTGAACATGAAGGGCTGATTTCTCTGAGTCCTTTAACTGAGATCGAAATGACAGCGTTTGAACCGGAGACATTGCAAAAACCGATCGTGTTACAGCCTTTGGATTGTACTGTTTCTGTCCAGCAGCCAGAAAAAGCCGTTTCCGAAGACAATAGAACATTGCTTGAAGAATTGAGGACAAGCCTGGCAGATATTTTTATTTTTGTCACCGGAAGATATTGATGCTGATGAGCCTTTCATCGACATGGGGCTTGACTCCATTATCGGGGGTTGAATGGATACAATCAGTAAATAAAACGTACCACACTGAGGTTACTGCCAATAAGGTGTATGAACACCCGACGCTTGAAGAGCTTGCGGAATATATAGCGGGACAAATAAAGAAACCCGAGTATGCTGAAGCGGAGGAAACAGCTGTACGTCCGCAAATCCCGGAAGCAGAAACTCCCCTTGCCGGCATCTCTGACGGCACTTTGCGCAAAGAGCTCAAAGATAGCCTTGCGGACATCCTGTTTCTGAAGCCGGAAGATATTGATGAGCATGAGGCATTCATTGAGATGGGGCTGGATTCTATTATCGGTGTGGAATGGGTTCAGTCCATTAATAAAAACCTATCAAGCGTCTATTACGGCAAATCTGGTCTATGAATATCCGACAATCGCAACCATTAGCCGGGTATCTGACCGGTTCTGTCCAAAAAGACACCGAGGAAGTACATGGGGAACATACGAAACAAAACGATATATCAGAAGAGACAGATGATCAAGAAGGCAACCGCCCGAGCTGATCTGTCTTAACCCGGGTGGAAAGGGAAGACCGATTTTCTGGTTTCATTCCGCGCTGGGGGCAGTGGAGCTGTACCAGCCTGTTGCCGAACGGACCGGGCGTCCTTTTTACGGAATTCAGGCAAGAGGGTGGATGACAGACCGGGAACCGATCCGCGGCATTAAAAACATGGCCGAATATTACGTTAAGCTCATACGTACCGTGCAGCCCCGGGGTCCTTACGATGTAGGCGGCTATTCATTAGGGGGCATGCTCGCTTATGAAGTCACCCGGCAATTGCAGCTGGCTGGAGAAACCGTTGAGTCAATCGTCATGCTCGACTCCTTTGACTCGGCGGGTGCGCGGGAAATAAACATTCCTGATGAATTATTTGTTAAAAAATATGATGCTGCAGACATGGCTTTATTTGCACCAATCTATCAACATCCGGAACGCATTTCTGAAACCTTGATTCATCGTGATGAACTGAATGCCGATCTTCATGAAGACATATTTTTAGAGCAATTAATCGCAAAAGCCAAACAGCGGGGGTTAACGAAGACGGACGCGTTGGTGAAAAATCTGATCAGGCAGAATATAAACGTTCAGGACGCTTACGAAATCACTGATTACGAACTCCTGCCGCTTCCGGATCCGAACGGGGTGACATGCTATTATTTCCGAAATAAAAACGGTATTTTCTTGGGTGAGCTTGATCCGTATTTTACGATCAGGGCTGATATTTTATCCGCTGATTTCGTCAACTATTGGAGCCAGTGGGAAGAGCATATGACAAACATCGCCATCATGGATGTGGCGGTATCCAACCATATGACGTTATTGGCAGAAGACAGCACAGCCGAAACGATTTCTGAATTTTGCGGCGTTTTATATTCGGCAGACGGAATGACAGAATACAGCTTTCTGCAGCTGCAGGAAAAGATACTGAATGTGCATGGTGAGATGCTGAAGCAGCCCGCCAAAACAAGACCGTAA
pUC57_*difL*_ACP_ACP_ACP_TE	AATGACCAGGTGTCATTAAAGGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACCGGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAAAGAGAAGCTGCTGGAACGGCTTTCAAGCAGTAGAGAATACGCCTTCTGTTTCACAGGATGAGCGTATCAACAACAGCAGCGATCATCACATATTAGCAAATGTACTAGGGATGGCTCGGATGAACTGCAGTTTCATAAGCCATTGCAGCAGTATGGATTTGATTCAATTTCTTGCATACAGTTATTACAGCAATTGCAATCAAAGGTGGACCCTCTCATTGTCTTGACGGAGCTTCAAGCATGCCATACTGTTCAGGACATGATGGACTTGATCGCAAAGAAACAGGAGGATACAGCGGCTGATGTAAACAGTTCTTTACCAGGCCAAGCCATGAAAAAGTCTGAAACGCAAACCGTTCTCCCGTCTTCGCCGCAGCCTGAACATGAAGGGCTGATTTCTCTGAGTCCTTTAACTGAGATCGAAATGACAGCGTTTGAACCGGAGACATTGCAAAAACCGATCGTGTTACAGCCTTTGGATTGTACTGTTTCTGTCCAGCAGCCAGAAAAAGCCGTTTCCGAAGACAATAGAACATTGCTTGAAGAATTGAGGACAAGCCTGGCAGATATTTTTATTTTTGTCACCGGAAGATATTGATGCTGATGAGCCTTTCATCGACATGGGGCTTGACTCCATTATCGGGGGTTGAATGGATACAATCAGTAAATAAAACGTACCACACTGAGGTTACTGCCAATAAGGTGTATGAACACCCGACGCTTGAAGAGCTTGCGGAATATATAGCGGGACAAATAAAGAAACCCGAGTATGCTGAAGCGGAGGAAACAGCTGTACGTCCGCAAATCCCGGAAGCAGAAACTCCCCTTGCCGGCATCTCTGACGGCACTTTGCGCAAAGAGCTCAAAGATAGCCTTGCGGACATCCTGTTTCTGAAGCCGGAAGATATTGATGAGCATGAGGCATTCATTGAGATGGGGCTGGATTCTATTATCGGTGTGGAATGGGTTCAGTCCATTAATAAAAACCTATCAAGCGTCTATTACGGCAAATCTGGTCTATGAATATCCGACAATCGCAACCATTAGCCGGGTATCTGACCGGTTCTGTCCAAAAAGACACCGAGGAAGTACATGGGGAACATACGAAACAAAACGATATATCAGAAGAGACAGATGATCAAGAACCGCCCGAGCTGATCTGTCTTAACCCGGGTGGAAAGGGAAGACCGATTTTCTGGTTTCATTCCGCGCTGGGGGCAGTGGAGCTGTACCAGCCTGTTGCCGAACGGACCGGGCGTCCTTTTTACGGAATTCAGGCAAGAGGGTGGATGACAGACCGGGAACCGATCCGCGGCATTAAAAACATGGCCGAATATTACGTTAAGCTCATACGTACCGTGCAGCCCCGGGGTCCTTACGATGTAGGCGGCTATTCATTAGGGGGCATGCTCGCTTATGAAGTCACCCGGCAATTGCAGCTGGCTGGAGAAACCGTTGAGTCAATCGTCATGCTCGACTCCTTTGACTCGGCGGGTGCGCGGGAAATAAACATTCCTGATGAATTATTTGATAAAAATATGATGCTGCAGACGGTGAACATGGCTTTATTTGCACCAATCTATCAACATCCGGAACGCATTTCTGAAACCTTGATTCATCGTGATGAACTGAATGCCGATCTTCATGAAGACATATTTTTAGAGCAATTAATCGCAAAAGCCAAACAGCGGGGGTTAACGAAGACGGACGCGTTGGTGAAAAATCTGATCAGGCAGAATATAAACGTTCAGGACGCTTACGAAATCACTGATTACGAACTCCTGCCGCTTCCGGATCCGAACGGGGTGACATGCTATTATTTCCGAAATAAAAACGGTATTTTCTTGGGTGAGCTTGATCCGTATTTTACGATCAGGGCTGATATTTTATCCGCTGATTTCGTCAACTATTGGAGCCAGTGGGAAGAGCATATGACAAACATCGCCATCATGGATGTGGCGGTATCCAACCATATGACGTTATTGGCAGAAGACAGCACAGCCGAAACGATTTCTGAATTTTGCGGCGTTTTATATTCGGCAGACGGAATGACAGAATACAGCTTTCTGCAGCTGCAGGAAAAGATACTGAATGTGCATGGTGAGATGCTGAAGCAGCCCGCCAAAACAAGACCGTAA
pUC57_*difL*_link_ACP_ACP_TE	CTGCTCTTGATACATGCATGGAAGGCGATGGAAGATGCAGGCTTAACAGGGCAGGTTTTATCCAGCCGCCCGACAGGAGTATTTGTCGCAGCCGGCAATACGGATACAGC

	TGTGGTTCCTTCCCTAATTCCAAACCGTATATCCTATGCACTTGATGTAAAAGGGCCAAGTGAATATTATGAAGCTGCCTGTTCCTCAGCTCTAGTGGCTTTGCACAGAGC TATACAATCCATTCGAAACGGCGAATGTGAGCAAGCCATTGTCGGGGCTGTGAATTTGCTGCTTTCACCAAAAGGCTTTATTGGCTTCGACTCAATGGGCTATTTGAGTGA GAAAGGGCAGGCCAAATCCTTTCAAGCAGATGCAAATGGCTTTGTCAGAAGTGAAGGAGCAGGAGTTCTCATCATTAAACCATTGCAAAAAGCCATTGAAGATTCTGAT CATATTTATTCGGTTATTAAAAGGTTCAGGTGTATCGCATGGCGGCAGGGGAATGTCACTTCACGCGCCAAATCCGGCCGGCATGAAGGATGCAATGCTGAAGGCTTATCA AGGAGCGCAAATTGATCCCAAAACGGTGACCTATATAGAAGCGCATGGGATCGCCTCTCCATTGGCAGACGCGATAGAAATAGAGGCGTTAAAGTCAGGCTGCAGTCAG CTCGAATTGGAACTTCCACAGGAAGTACGGGAGGAAGCGCCATGTTATATCAGCAGCTTAAAGCCGAGCATCGGACACGGTGAACTCGTCTCAGGCATGGCTGCTCTTAT GAAGGTCAGCATGGCGATGAAGCATCAAACAATACCAGGCATATCCGGATTTTCGTCTTTGAATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGA ATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGA AGCGCCGCAAACCGTGCGCCGGTCGCCGGCAAAACCAGTCTGCCTCATTCCTTTATCTGCGCGGACGGAGCCGTCAGCTGGAAATTCAGGCAGAACGCCTGCTGGATTCAC TCAGCAAGAACAAAGACAGCGATCTGGGAAATATAAGTTATACGCTCCTGCTCGGCAGAAAGCATCTACATCACCGGTTTGCCTGTATCGTAAGCAGTGCGGACGAACTG AAGCGGGTATTATCGGAGTGGCTTGTAAAAAGAGAGCTTCCGGGTGTGTTTGTGTCCAATTTGAAGGATCAGAAGCCGGCAGAAGAAGCCGGAATGAAAACATTCGGAA TGGAATGCATTGAACAGTGCCGCTCGGCGGCTTCGCCCGCACAGTACAGAGAAAACCTGGCACATATCGCTGACCTTTTTTGCAGAGGGTATGATCTTCCTTTTACACGGC TTTTTGACGGAGGACAATATTATAAAACGCCGCTGCCTGCCTACCCGTTTTTAAAAGAAAGGTATTGGGCGGCTGATGTAAACAGTTCTTTACCAGGCCAAGCCATGAAA AAGTCTGAAACGCAAACCGTTCTCCCGTCTTCGCCGCAGCCTGAACATGAAGGGCTGATTTCTCTGAGTCCTTTAACTGAGATCGAAATGACAGCGTTTGAACCGGAGAC ATTGCAAAAACCGATCGTGTTACAGCCTTTGGATTGTACTGTTTCTGTCCAGCAGCCAGAAAAAGCCGTTTCCGAAGACAATAGAACATTGCTTGAAGAATTGAGGACAA GCCTGGCAGATATTTTATTTTTGTCACCGGAAGATATTGATGCTGATGAGCCTTTCATCGACATGGGGCTTGACTCCATTATCGGGGTTGAATGGATACAATCAGTAAATA AAACGTACCACACTGAGGTTACTGCCAATAAGGTGTATGAACACCCGACGCTTGAAGAGCTTGCGGAATATATAGCGGGACAAATAAAGAAACCCGAGTATGCTGAAGC GGAGGAAACAGCTGTACGTCCGCAAATCCCGGAAGCAGAAACTCCCCTTGCCGGCATCTCTGACGGCACTTTGCGCAAAGAGCTCAAAGATAGCCTTGCGGACATCCTGT TTCTGAAGCCGGAAGATATTGATGAGCATGAGGCATTCATTGAGATGGGGCTGGATTCTATTATCGGTGTGGAATGGGTTCAGTCCATTAATAAAACCTATCAAGCGTCT ATTACGGCAAATCTGGTCTATGAATATCCGACAATCGCAACATTAGCCGGGTATCTGACCGGTTCTGTCCAAAAAGACACCGAGGAAGTACATGGGGAACATACGAAAC AAAACGATATATCAGAAGAGACAGATGATCAAGAACCGCCCGAGCTGATCTGTCTTAACCCGGGTGGAAAGGGAAGACCGATTTTCTGGTTTCATTCCGCGCTGGGGGGC AGTGGAGCTGTACCAGCCTGTTGCCGAACGGACCGGGCGTCCTTTTTACGGAATTCAGGCAAGAGGGTGGATGACAGACCGGGAACCGATCCGCGGCATTAAAAACATG GCCGAATATTACGTTAAGCTCATACGTACCGTGCAGCCCCGGGGTCCTTACGATGTAGGCGGCATGCTCGGTTTGATCTGTATGTGAAGTCACCCGGCAATTGCA GCTGGCTGGAGAAACCGTTGAGTCAATCGTCATGCTCGACTCCTTTGACTCGGCGGGTGCGCGGGAAATAAACATTCCTGATGAATTATTTGATAAAAATATGATGCTGC AGACGGTGAACATGGCTTTATTTGCACCAATCTATCAACATCCGGAACGCATTTCTGAAACCTTGATTCATCGTGATGAACTGAATGCCGATCTTCATGAAGACATATTTT TAGAGCAATTAATCGCAAAAGCCAAACAGCGGGGGGTTAACGAAGACGGACGCGTTGGTGAAAAATCTGATCAGGCAGAATATAAACGTTCAGGACGCTTACGAAATCAC TGATTACGAACTCCTGCCGCTTCCGGATCCGAACGGGGTGACATGCTATTATTTCCGAAATAAAAACGGTATTTTCTTGGGTGAGCTTGATCCGTATTTTACGATCAGGGC TGATATTTTATCCGCTGATTTCGTCAACTATTGGAGCCAGTGGGAAGAGCATATGACAAACATCGCCATCATGGATGTGGCGGTATCCAACCATATGACGTTATTGGCAG AAGACAGCACAGCCGAAACGATTTCTGAATTTTGCGGCGTTTTATATTCGGCAGACGGAATGACAGAATACAGCTTTCTGCAGCTGCAGGAAAAGATACTGAATGTGCAT GGTGAGATGCTGAAGCAGCCCGCCAAAACAAGACCGTAA
pUC57_*misF*_TE	AATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTA TCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTC TTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACC GGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTT CTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAAC CTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGTCTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAA AGAGAAGCTGCTGGAACGGCTTTCAAGCAGTAGAGAATACGCCTTCTGTTTCACAGGATGAGCGTATCAACAACAGCAGCGATCATCACATATTAGCAAATGTACTAGG GATGGCTCCGGATGAACTGCAGTTTCATAAGCCATTGCAGCAGTATGGATTTGATTCAATTTCTTGCATACAGTTATTACAGCAATTGCAATCAAAGGTGGACCCTCTCAT TGTCTTGACGGAGCTTCAAGCATGCCATACTGTTCAGGACATGATGGACTTGATCGCAAAGAAACAGGAGGATACACAAACCGAGCCGGAGCTGGTGAACCTTTCCGGA GCCGGTCTTTCTTCATCACATGACACACGTTTTCCAGACTGGTCCATATGAACTCAAGTACCAGGGGAAGACCTGTATTCTGGTTCCATGGCATCGCGGGCGTTACGGTT TACGAGCCTGTTGCTGAAAAAAGCCAGCGCCCCTTCTATGGCGTTCAGCCCTGTAGCTGGATCAACCAAACCGCGGGACCTTCTCATATTCAGGCCATGGTGCAAGGTTA TCTGGATGCCATACGATCCGTGCAGCCGGAAGGTCCTTACGATTTTGGTGGCTATTCGCTGGGTGGCATGTTGGCTTATGAAGCCACCCGTCAATTACAGGACATGGGAG AGAGCGTATCGACCATTGTGATGGTCGATACGCTACAGAGTAATCCTGATCGTAAGGATAAGTACTCGCGCAAAACAGATTATCTGGTTGTTTTGAATCGGTCCTTTGCCT CATCTGCATGGCAGAGCTCGGAAAATACCGTGCAGATCACTGTCGTTGCAGATGAGCTGGATGCAAGTCTAAGTGACGACGAGTATTTAACAGAGCTGATTGTCTTA GCCAAACAACGAGGACTGGTCAAGACAGAGACAGAGATCCGTGCCAGCCTTGAATATGCAACAACCCTTGATGAGTTTTTATCAAACTGATCCCTTTACGGTTAGGCCCCT ATCTGATCCCGATGCCGTTAACTGCTACTATTTCCGCAATAAGAATGGCGCGATGTTGGGTGATCAGGCGCCCTACTATTTTGCGACACCACAAGACAGGGAAATATTTGT CACCACCGATCAGTCAGCCGATTCGCGGCAGTGGGAAAAGAACTTGACGAACCTTCACATCATTGATATCGATTCGTCCAATCACATGACGATGTTCTCGGAAGAAAAAC CACGCCAGACCATTCTTGAGTTCTGTGAAATCTTGTATTCTGAAGCCGGACTTCCGAATGCATTTCTTGACTTCTTTATGGCCAAGGCTCGAGAAATACATGGCAATATAG AGCTGGACAGATAA
pUC57_*misF*_ACP_TE	CCAAATCCGGCCGGCATGAAGGATGCAATGCTGAAGGCTTATCAAGGAGCGCAAATTGATCCCAAAACGGTGACCTATATAGAAGCGCATGGGATCGCCTCTCCATTGG CAGACGCGATAGAAATAGAGGCGTTAAAGTCAGGCTGCAGTCAGCTCGAATTGGAACTTCCACAGGAAGTACGGGAGGAAGCGCCATGTTATATCAGCAGCTTAAAGCC GAGCATCGGACACGGTGAACTCGTCTCAGGCATGGCTGCTCTTATGAAGGTCAGCATGGCGATGAAGCATCAAACAATACCAGGCATATCCGGATTTTCGTCTTTGAATG ACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAA CAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTCTTTC TGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACCGGCC

	GAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTTCTGT ACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAACCTTG AAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAAAGAG AAGCTGCTGGGTTGAGAGTAAAGCCCTGTCTCAAGTTGTAACCACGGCTGAGGCCGAATTACTTTCGAAAAGTACGGATGATCATATTTCCGTTGACAGAAATGAAGAAT CAGTTCTTTCAGAGATTCTGTGCCGAACCTTAGGCCTAACAGCAGAAGAGTTTAATCAAAAGACCCCATTGGTTCAATATGGGGTGGATTCCGTTATGTTTGTGCAAATTT TTCAACAAATTAAAGACAAGATTGATAACCGCGTTAACTTGAGTCAGCTGCTGGAATGCCAAACCATGCAAGATATGATTTTCTATTTGACTGCGAAAAAGGAGGACGCG CAAACCGAGCCGGAGCTGGTGAACCTTTCCGGAGCCGGTCTTTCTTCATCACATGACACACGTTTTCCAGAGCTGGTCCATATGAACTCAAGTACCAGGGGAAGACCTGT ATTCTGGTTCCATGGCATCGCGGGCGTTACGGTTTACGAGCCTGTTGCTGAAAAAAGCCAGCGCCCCTTCTATGGCGTTCAGCCCTGTAGCTGGATCAACCAAACCGCGG GACCTTCTCATATTCAGGCCATGGTGCAAGGTTATCTGGATGCCATACGATCCGTGCAGCCGGAAGGTCCTTACGATTTTGGTGGCTATTCGCTGGGTGGCATGTTGGCTT ATGAAGCCACCCGTCAATTACAGGACATGGGAGGAGCGTATCGACCATTGTGATGGTCGATACGCTACAGAGTAATCCTGATCGTAAGGATAAGTACTCGCGCAAAAC AGATTATCTGGTTGTTTTGAATCGGTCCTTTGCCTCATCTGCATGGCAGAGCTCGGAAAATACCGTGCAGGACATGCTGATTCGTGCAGATGAGCTGGATGCAAGTCTAAG TGACGACGAGTATTTAACAGAGCTGATTGTCTTAGCCAAACAACGAGGACTGGTCAAGACAGAGACAATCCGTGCCAGCCTTGAATATGCAACAACCCTTGATGAGTTTT ATCAAACTGATCCCTTTACGGTTAGGCCCCTATCTGATCCCGATGCCGTTAACTGCTACTATTTCCGCAATAAGAATGGCGCGATGTTGGGTGATCAGGCGCCCTACTATT TTGCGACACCACAAGACAGGGAAATATTTGTCACCACCGATCAGTCAGCCGATTCGCGGCAGTGGGAAAGAACTTGACGAACCTTCACATCATTGATATCGATTCGTCC AATCACATGACGATGTTCTCGGAAGAAAAACCACGCCAGACCATTCTTGAGTTCTGTGAAATCTTGTATTCTGAAGCCGGACTTCCGAATGCATTTCTTGACTTCTTTATG GCCAAGGCTCGAGAAATACATGGCAATATAGAGCTGGACAGATAA
pUC57_*misF*_ACP_ACP_TE	AATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTA TCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCACATTGTAGTTC TTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATACACTTCAAACC GGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAGCTGAAGAGTT CTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCTGAAAACAAC CTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATCCATTTGAAA AGAGAAGCTGCTGGGTTGAGAGTAAAGCCCTGTCTCAAGTTGTAACCACGGCTTTCAACAGCAGCTGTTTCACAGGATGAGCGTATCAACAACAGCGGATCATCACATATTAGCAAATGTACTAGG GATGGCTCCGGATGAACTGCAGTTTCATAAGCCATTGCAGCAGTATGGATTTGATTCAATTTCTTGCATACAGTTATTACAGCAATTGCAATCAAAGGTGGACCCTCTCAT TGTCTTGACGGAGCTTCAAGCATGCCATACTGTTCAGGACATGATGGACTTGATCGCAAAGAAACAGGAGGATACAGTTGAGAGTAAAGCCCTGTCTCAAGTTGTAACCA CGGCTGAGGCCGAATTACTTTCGAAAAGTACGGATGATCATATTTCCGTTGACAGAAATGAAGAATCAGTTCTTTCAGAGATTCTGTGCCGAACCTTAGGCCTAACAGCA GAAGAGTTTAATCAAAAGACCCCATTGGTTCAATATGGGGTGGATTCCGTTATGTTTGTGCAAATTTTTCAACAAATTAAAGACAAGATTGATAACCGCGTTAACTTGAGT CAGCTGCTGGAATGCCAAACCATGCAAGATATGATTTTCTATTTGACTGCGAAAAAGGAGGACGCGCAAACCGAGCCGGAGCTGGTGAACCTTTCCGGAGCCGGTCTTTC TTCATCACATGACACACGTTTTCCAGAGCTGGTCCATATGAACTCAAGTACCAGGGGAAGACCTGTATTCTGGTTCCATGGCATCGCGGGCGTTACGGTTTACGAGCCTGT TGCTGAAAAAAGCCAGCGCCCCTTCTATGGCGTTCAGCCCTGTAGCTGGATCAACCAAACCGCGGGACCTTCTCATATTCAGGCCATGGTGCAAGGTTATCTGGATGCCA TACGATCCGTGCAGCCGGAAGGTCCTTACGATTTTGGTGGCTATTCGCTGGGTGGCATGTTGGCTTATGAAGCCACCCGTCAATTACAGGACATGGGAGGAGCGTATCG ACCATTGTGATGGTCGATACGCTACAGAGTAATCCTGATCGTAAGGATAAGTACTCGCGCAAAACAGATTATCTGGTTGTTTTGAATCGGTCCTTTGCCTCATCTGCATGG CAGAGCTCGGAAAATACCGTGCAGGACATGCTGATTCGTGCAGATGAGCTGGATGCAAGTCTAAGTGACGACGAGTATTTAACAGAGCTGATTGTCTTAGCCAAACAAC GAGGACTGGTCAAGACAGAGACAGAGATCCGTGCCAGCCTTGAATATGCAACAACCCTTGATGAGTTTTATCAAACTGATCCCTTTACGGTTAGGCCCCTATCTGATCCC GATGCCGTTAACTGCTACTATTTCCGCAATAAGAATGGCGCGATGTTGGGTGATCAGGCGCCCTACTATTTTGCGACACCACAAGACAGGGAAATATTTGTCACCACCGA TCAGTCAGCCGATTCGCGGCAGTGGGAAAGAACTTGACGAACCTTCACATCATTGATATCGATTCGTCCAATCACATGACGATGTTCTCGGAAGAAAAACCACGCCAGA CCATTCTTGAGTTCTGTGAAATCTTGTATTCTGAAGCCGGACTTCCGAATGCATTTCTTGACTTCTTTATGGCCAAGGCTCGAGAAATACATGGCAATATAGAGCTGGACA GATAA
pUC57_*misF*_link_ACP_TE	CTGCTCTTGATACATGCATGGAAGGCGATGGAAGATGCAGGCTTAACAGGGCAGGTTTTATCCAGCCGCCCGACAGGAGTATTTGTCGCAGCCGGCAATACGGATACAGC TGTGGTTCCTTCCCTAATTCCAAACCGTATATCCTATGCACTTGATGTAAAAGGGCCAAGTGAATATTATGAAGCTGCCTGTTCCTCAGCTCTAGTGGCTTTGCACAGAGC TATACAATCCATTCGAAACGGCGAATGTGAGCAAGCCATTGTCGGGGCGTGTGAATTTGCTGCTTTCACCAAAAGGCTTTATTGGCTTCGACTCAATGGGCTATTTGAGTGA GAAAGGGCAGGCCAAATCCTTTCAAGCAGATGCAAATGGCTTTGTCAGAAGTGAAGGAGCAGGAGTTCTCATCATTAAACCATTGCAAAAAGCCATTGAAGATTCTGAT CATATTTATTCGGTTATTAAAGGTTCAGGTGTATCGCATGGCGGCAGGGGAATGTCACTTCACGCGCCAAATCCGGCCGGCATGAAGGATGCAATGCTGAAGGCTTATCA AGGAGCGCAAATTGATCCCAAAACGGTGACCTATATAGAAGCGCAATGGAATCGCCCTCTCCATTGGCAGACGCGATAGAAATAGAGGCGTTAAAGTCAGGCTGCAGTCAG CTCGAATTGGAACTTCCACAGGAAGTACGGGAGGAAGCGCCATGTTATATCAGCAGCTTAAAGCCGAGCATCGGACACGGTGAACTCGTCTCAGGCATGGCTGCTCTTAT GAAGGTCAGCATGGCGATGAAGCATCAAACAATACCAGGCATATCCGGATTTTCGTCTTTGAATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGA ATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCAGAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGA ATATATTGCTTCTGCGGGTAAGAACGCCGTGTCACCTGATATAGAGTTGGCAGCCGATCGCGAGTTGATGGTTTTTTCCGCAAAAAGCTGTGATCGTCTTCAGGCGATGCT GCAACAGATGCGTCATTATGTCGAATCCAACCAAGCACTTTCGCTAAAGAAGATGGCCTATACGCTCCAGGTAGGACGTATGGCGATGACACATCGAATCGCGATGCTGG TGCATAATCGAGAAGAGCTGATCCAGGGCATCCAAGATGCTTTGCATAGCCTAGAACAACCAGCCGGGATAATCAACACAACCATTCCAATATACATGGGTGATCTGGA GGCGGATCACTCGGAAGTAAAGCATTTGCTCTCCGGTCAATCCGGAGAGGCCATGCTGAAAGTGCTTTTCGCCGAAGAAAATCTGGAAAAAATTGCTCTTTACTGGGTTA AAGGAGGCCATATTCCCTGGCTGGCACTGCATCAAAAAGGTCAAATGTCTATGTTGTCTCTACCCACGTACCCTTTCTCCAAGCGGTACTGCTGGGTTGAGAGTAAAGCCC TGTCTCAAGTTGTAACCACGGCTGAGGCCGAATTACTTTCGAAAAGTACGGATGATCATATTTCCGTTGACAGAAATGAAGAATCAGTTCTTTCAGAGATTCTGTGCCGAA CCTTAGGCCTAACAGCAGAAGAGTTTAATCAAAAGACCCCATTGGTTCAATATGGGGTGGATTCCGTTATGTTTGTGCAAATTTTTCAACAAATTAAAGACAAGATTGAT AACCGCGTTAACTTGAGTCAGCTGCTGGAATGCCAAACCATGCAAGATATGATTTTCTATTTGACTGCGAAAAAGGAGGACGCGCAAACCGAGCCGGAGCTGGTGAACC

	TTTCCGGAGCCGGTCTTTCTTCATCACATGACACACGTTTTCCAGAGCTGGTCCATATGAACTCAAGTACCAGGGGAAGACCTGTATTCTGGTTCCATGGCATCGCGGGCG TTACGGTTTACGAGCCTGTTGCTGAAAAAAGCCAGCGCCCCTTCTATGGCGTTCAGCCCTGTAGCTGGATCAACCAAACCGCGGGACCTTCTCATATTCAGGCCATGGTGC AAGGTTATCTGGATGCCATACGATCCGTGCAGCCGGAAGGTCCTTACGATTTTGGTGGCTATTCGCTGGGTGGCATGTTGGCTTATGAAGCCACCCGTCAATTACAGGACA TGGGAGAGAGCGTATCGACCATTGTGATGGTCGATACGCTACAGAGTAATCCTGATCGTAAGGATAAGTACTCGCGCAAAACAGATTATCTGGTTGTTTTGAATCGGTCC TTTGCCTCATCTGCATGGCAGAGCTCGGAAAAATACCGTGCAGGACATGCTGATTCGTGCAGATGAGCTGGATGCAAGTCTAAGTGACGACGAGTATTTAACAGAGCTGAT TGTCTTAGCCAAACAACGAGGACTGGTCAAGACAGAGACAGAGATCCGTGCCAGCCTTGAATATGCAACAACCCTTGATGAGTTTTATCAAACTGATCCCTTTACGGTTA GGCCCCTATCTGATCCCGATGCCGTTAACTGCTACTATTTCCGCAATAAGAATGGCGCGATGTTGGGTGATCAGGCGCCCTACTATTTTGCGACACCACAAGACAGGGAA ATATTTGTCACCACCGATCAGTCAGCCGATTCGCGGCAGTGGGAAAAGAACTTGACGAACCTTCACATCATTGATATCGATTCGTCCAATCACATGACGATGTTCTCGGA AGAAAAACCACGCCAGACCATTCTTGAGTTCTGTGAAATCTTGTATTCTGAAGCCGGACTTCCGAATGCATTTCTTGACTTCTTTATGGCCAAGGCTCGAGAAATACATGG CAATATAGAGCTGGACAGATAA
pUC57_*misF*_KS^0_ACP_TE	GTGCTTGTCGGTGCCACATATGGCAGTCTGGCCATTGAAGCATTTTTTAACCTGTTCCCTGAGGAAAACAGCGGCCGTATCAGCAAATTAAGTTATATCAGTCCAATTGTC ATCAAACAAGGCGAGACCATTGAACTTCAGGCAAAGCCGCTGCAAAAAGATCAAGTCATAGAACTGCAAATCATGTATCGCGAGCCGTCCTCTGGTTTGTGGAAGCCTG CCGCAATCGGACAATGCGGAATCGGTTCTTTTGAGCCCAAAAAAGTCAATATCGAGAACGTTAAGCATTCATTAACTAAGCTTCATCACATCGATCAGATGTATAAAACC GGAAACGGTCCTGAATGGGGAGAGTTATTTAAGACAATTACTCATCTCTACAGAGATCACAAGTCTATACTGGCAAAAATTCGCCTGCCCCAAAGCGGGCTGGCAAACG GGCACCATTACACTGTAAGCCCATTGATGACAAACAGCGCGTACTTGGCTATCCTCAGTTTCTTAGAGCAGTTTGACATGACAGGCGGCTTCCTGCCGTTTGGAATCAATG ATATCCAGTTTACAAAGCAAACGATAAAAGGGGATTGCTGGCTTTTGATTACATTGGTTAAGAATACAGGTGACATGTTGCTGTTTGATGTAGATGTGATCAATGAGTCGT CAGAAACAGTGCTGCACTACTCGGGCTACTCGCTTAAACAGCTTCGTATTTCGAATCAAAGAGGAAATCAAAATAAGGCCATCAAAGCCAGCAATCTGAAAGCTCGTATC AGAAGCTATGTAACAGATAAACTGGCAGTAAACATGGCCGATCCGTCAAAATTGTCAATTGCAAAAGCGCATATCATGGATTTTGGAATTGATTCTTCTCAATTGGTTGC ACTGACAAGGGAGATGGAAGCAGAGACAAAGATCGAATTAAATCCGACTCTGTTTTTTGAATATCCGACTATTCAAGAGTTAATCGACTTTTTTGCGGACAAACATGAAG CATCTCTCCACGTGGTCCTTGGCTTAGACAAGTCAAAGGCACCTGAAACTGACCACTCGGCTGTTGTCGATCCGGCTCGCCCAAAGGTACATCGGCGTTCTCGCTTGTTGG CACCGACATTAGATGAATCACCCACTCCCCGTTGCGCAAACGCAGGATGTCGCCATTATCGGTGTTGCCGGACGTTATCCCCAATCCCCGGATCTTGAATCGTTCTGGGAAA ATTTGTCGCAAGGTAAAAATTGTATTACTGAAGTTCCCGAAAGTCGTTGGGACCCCAGCTTGTATGTTGATTTTGACCCTTCCAACTCTGGTCAGGTGTATAGCAAGTGGG GTGGTTTCTTAGATGACGTGGACAAATTTGACGCTTTATTTTTCAATATTTCTCCTCGTGAAGCAGAGTTGATGAACCCGAATGAGCGTCTTCCTGGAAACCGTCTGGG AGCTGGTGGAACGTTCGGGGTATACACCGGAAAGAATACAGGAACAGCACGAAAATAACGTGGGGGTTTATATCGGGGGCCATGTATCAGCAGTATCAGGCCTTAGCTTC CGAGCCCCTCATAGAGTCGGTCCTGTCCTTGTCTTCCCATAGTGCGATTGCGAATCGCGTGTCCTATTTTTTTGATTTCCAGGGCCCTAGTCTGGCCATCGACACGATGTGT TCCTCTTCATTAATTGCGATACAGTTAGCTTGTGACAGCCTGCTGAAAGGCGAGTGCCAGATGGCCATTGCCGGAGGTGTCAACCTCACGATTCATCCGAACAAATATATT GGCTTGAGCCTGATCAAATGATCAGTCGGCAAAACACCAGCAAAAGGTTTGGTGAGGGTGATGGTTATCTTCCGGCAGAAGGTGTCGGTGCCGTGTTGTTAAAGCCCTT GGCCAAGGCGATCGAAGATCAGGATTCCATTTTAGCCGTCATTAAATCCATTAGCACCAATCATAGCGGACATACGTATGAGTTTCGTGTACCCAACCCGAATGCGCAAG CCCAATTATTTGAAGCAAATTTCCGTAAATCCGGCATCGATCCACGCACGATTACGTATGTTGAATCGGCGGCCAATGGTTCTCCCTTGGGAGATCCGATCGAGGTCAGTG CACTTCGTAAAGCGTTTCAAAAATTCACTGCGGATCAGGGCTTTTGTGCGATTGGTTCCGTCAAATCCAATATCGGTCATGCAGAGGCAGCATCCGGCATCTCCCAACTGA CAAAAGTGATATTACAGCTTCAGCATCAGCAACTTGTTCCGTCGATTAATGCAGACCCACTGAATCCAAATTTAAGCTTCGATGGTTCGCCATTTTATCTGCAACAACAAC TGCAACCATGGGCACGTCCAGTCTTCAATATCGATGGTCAAGAGCGTGAGTTCCCCGGCGCGCAACCGTTAGCTCCTTTGGGGCGGGAGGCTCGAATGCGCATTTGATT ATCGAGGAATATATTGCTTCTGCGGGTAAGAACGCCGTGTCACCTGATATAGAGTTGGCAGCCGATCGCGAGTTGATGGTTTTTTCCGCAAAAAGCTGTGATCGTCTTCAG GCGATGCTGCAACAGATGCGTCATTATGTCGAATCCAACCAAGCACTTTCGCTAAAGAAGATGGCCTATACGCTCCAGGTAGGACGTATGGCGATGACACATCGAATCGC GATGCTGGTGCATAATCGAGAAGAGCTGATCCAGGGCATCCAAGATGCTTTGCATAGCCTAGAACAACCAGCCGGGATAATCAACACAACCATTCCAATATACATGGGT GATCTGGAGGCGGATCACTCGGAAGTAAAGCATTTGCTCTCCGGTCAATCCGGAGAGGCCATGCTGAAAGTGCTTTTCGCCGAAGAAAATCTGGAAAAAATTGCTCTTTA CTGGGTTAAAAGGAGGCCATATTCCCTGGCTGGCACTGCATCAAAAAGGTCAAATGTCTATGTTGTCTCTACCCACGTACCCTTTCTCCAAGCGGTACTGCTGGGTTGAGAG TAAAGCCCTGTCTCAAGTTGTAACCACGGCTGAGGCCGAATTACTTTCGAAAAGTACGGATGATCATATTTCCGTTGACAGAAATGAAGAATCAGTTCTTTCAGAGATTCT GTGCCGAACCTTAGGCCTAACAGCAGAAGAGTTTAATCAAAAGACCCCATTGGTTCAATATGGGGTGGATTCCGTTATGTTTGTGCAAATTTTTCAACAAATTAAAGACA AGATTGATAACCGCGTTAACTTGAGTCAGCTGCTGAAGAACCATGCAAGATATGATTTTCTATTTGACTGCGAAAAAGGAGGACGCGCAAACCGAGCCGGAGCT GGTGAACCTTTCCGGAGCCGGTCTTTCTTCATCACATGACACACGTTTTCCAGAGCTGGTCCATATGAACTCAAGTACCAGGGGAAGACCTGTATTCTGGTTCCATGGCAT CGCGGGCGTTACGGTTTACGAGCCTGTTGCTGAAAAAAGCCAGCGCCCCTTCTATGGCGTTCAGCCCTGTAGCTGGATCAACCAAACCGCGGGACCTTCTCATATTCAGG CCATGGTGCAAGGTTATCTGGATGCCATACGATCCGTGCAGCCGGAAGGTCCTTACGATTTTGGTGGCTATTCGCTGGGTGGCATGTTGGCTTATGAAGCCACCCGTCAAT TACAGGACATGGGAGAGAGCGTATCGACCATTGTGATGGTCGATACGCTACAGAGTAATCCTGATCGTAAGGATAAGTACTCGCGCAAAACAGATTATCTGGTTGTTTTG AATCGGTCCTTTGCCTCATCTGCATGGCAGAGCTCGGAAAAATACCGTGCAGGACATGCTGATTCGTGCAGATGAGCTGGATGCAAGTCTAAGTGACGACGAGTATTTAAC AGAGCTGATTGTCTTAGCCAAACAACGAGGACTGGTCAAGACAGAGACAGAGATCCGTGCCAGCCTTGAATATGCAACAACCCTTGATGAGTTTTATCAAACTGATCCCT TTACGGTTAGGCCCCTATCTGATCCCGATGCCGTTAACTGCTACTATTTCCGCAATAAGAATGGCGCGATGTTGGGTGATCAGGCGCCCTACTATTTTGCGACACCACAAG ACAGGGAAATATTTGTCACCACCGATCAGTCAGCCGATTCGCGGCAGTGGGAAAAGAACTTGACGAACCTTCACATCATTGATATCGATTCGTCCAATCACATGACGATG TTCTCGGAAGAAAAACCACGCCAGACCATTCTTGAGTTCTGTGAAATCTTGTATTCTGAAGCCGGACTTCCGAATGCATTTCTTGACTTCTTTATGGCCAAGGCTCGAGAA ATACATGGCAATATAGAGCTGGACAGATAA
pUC57_*pksS*_KO	CGCCCTGTCTTTTGGTTCCATGGCGGAGTAGGAGGAGTTGAAATCTATCAGCAATTTGCACAAAAGAGCCAGCGCCCTTTTTACGGCATTCAAGCCAGAGGATTCATGAC TGATTCTGCTCCTTTGCACGGAATTGAACAAATGGCTTCCTATTATATAGAGATCATTCGATCCATACAGCCTGAAGGTCCTTATGATGTAGGCGGATATTCCTTAGGCGG GATGATTGCATATGAAGTCACTCGCCAGCTGCAAAGCCAAGGCCTTGCTGTCAAAAGCATGGTGATGATTGACTCCCCATATCGTTCTGAGACAAAGAGAATGAGGCAT CTATGAAAACGTCAATGCTGCAAACAATTAATACGATGCTGGCATCGATTGCGAAACGGGAAAAGTTTACGGATGTTCTCATCAGCCGTGAAGAGGTGGACATAAGCTTA GAGGATGAAGAATTCCTGTCTGAGTTGATTGACTTGGCAAAAGAACGAGGGCTAAACAAACCAGATAAACAAATACGTGCGCAGGCTCAGCAAATGATGAAAACACAG

	CGCGCCTATGATTTGGAGTCGTACACTGTTAAGCCTCTCCCTGATCCTGAGACGGTAAAATGTTATTATTTCCGCAACAAAAGCAGGTCTTTCTTTGGTGATTTAGACACTT ATTTCACTTTATCAAATGAAAAAGAACCGTTTGATCAAGCTGCCTATTGGGAGGAATGGGAGCGGCAAATTCCTCATTTCCACCTGGTGGATGTCGATTCAAGCAACCAC TTCATGATATTAACCGAACCGAAAGCGTCAACAGCCCTGTTAGAATTTTGCGAAAAGCTCTATTCAAACAGGGGAGTAGTGAATGCGAATTTCCTTAAGGCTTTCCGGAA GAAACATGAAGCGAGGGAAGAAAAAGAAACAGATGAATTGGTGAAGCGCTGACTGGAGAGACCCTCGTTTAAACTACACAGCCCAGTCCAGACTATTCGGCACTGAAA TTATGGGTGAAGTGGTCAAGACCTCACTAGGCACCTTAAAAATAGCGCACCCTGAAGAAGATTTATTTGAGGTAGCCCTTGCCTACCTAGCTTCCAAGAAAGATATCCTA ACAGCACAAGAGCGGAAAGATGTTTTGTTCTACATCCAGAACAACCTCTGCTAAAATTCCTGAAAAATTTTGCAAAAAGTTGTTGACTTTATCTACAAGGTGTGGCATAA TGTGTGAATCCGGTTAAGCTTACATAAGGAGGAACTACTATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGC ATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCGGGCCTTTATTCA CATTCTTGCCCGCCTGATGAATGCTCATCCGGAGTTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCA AACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCC TAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACT ATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACA ACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAACGTATGCACGATCTGTTACGACCTTCAATCCTTCGTTTTCTTGGTGAATATGAAGCTCACCTAATAAAGGAATCTC ATAGGCATTTTGCGGAAGCGGTACGTTTCCGTCCTCAGTGAATGTTGTTCCTTCGCCTTCTAAAACAAGCTCTTCTTTCGCAACATAATCATCAGGATGATTCGTACTGCGT ATAGCCACTTTTTGAAGGTGGAGAACAATCTGATCCAAGTCTGATATCTCTTCATCATGGGTCAATTCACCTTTTCTAATATCAAGCGTCTGGCCGACTTTTGATTCCAGCC ATTGAGAAAGCTGTGTATTCGTGTGTGCCATCAATATTCCTCCTTTTGATCTACACGATACTATTCCCAATTGCTGCATCTTTTACACGGGAAAGAGCCGCCCCACTCTTTA TGTATGAAACAGTTCGCTGTTGAGTGCCTTGATATAGCGTACGGCAGAACGCTGGACCGATTCGTCGGCATTTTCTTTGACAACTTTAGAGAAAGCATTGTAAAGCCGGTT GAATTGAAGCGGTTTTACTTTGGCTGCTATTTCTTCAACTTTGGATGACTTTCCGGCAGTTTCTACACATATATTGGGTAGCTGTACATAAAGCTGACCCAGTTTCGATCCGCCGCAAC CGTAATGATATCTCCGGTCAATAGACAGCCTTTCCCTTCATTGCCCCTTGCTCCAATGCAAAACAGCACCGCCTTTAAAATGTCCGCCCAGTCGGTAAAGGTCCAGTCCCGG CTTCATATTGAGCGTTTCCCCTGACCAGAAGTGAATGTGATTGCTTGGCCTTGTCACCCATTCTTTATCATCCTCATGGATATAGATCGGCGCCTGAAACGCTTCAGCCCAC TCAACCTGAGTAGAGTAATAATGTGGATGAGACAAGGCGATGGCTTGAATACCGCCTAATTCGTTGATTTGGTCAATTGTCTTTTGGTCAAGATATGTCATGCAATCCCAC AGCACATTGAAGCCCTTATGCTGAATAAGATGGGCTGTTTGTCCAATCGCAAACTCTGGTTCAGT
pUC57_*lumE*_ACP_ACP_TE	ATGGCTAGCAATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCA GAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCAC ATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATAC ACTTCAAACCGGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAG CTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCT GAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATC CATTTGAAAAGAGAAGCTGCTGGAACGGCTTTCAAGCAGTAGAGAATACGCCTTCTGTTTCACAGGATGAGCGTATCAACAACAGCAGCGATCATCACATATTAGCAAAT GTACTAGGGATGGCTCCGGATGAACTGCAGTTTCATAAGCCATTGCAGCAGTATGGATTTGATTCAATTTCTTGCATACAGTTATTACAGCAATTGCAATCAAAGGTGGAC CCTCTCATTGTCTTGACGGAGCTTCAAGCATGCCATACTGTTCAGGACATGATGGACTTGATCGCAAAGGAGGCTAAAGTTGAGGTTAACGTTCCATCTCAAGTTGATCTC AAGGTTCAGACAGAGTCGGTTTCTAAAAAACCTTGATGATCAAAGCCCCTCTGGCAAAAACAGAGAGGCAGCACTTGTGGATATTATGCTTCTAATCTTAGGACTGACGGC AGAAGAATTAAATCTAAATACCCCTCTGGTTCAATATGGGGTTGATTCTATTATGTTTATGCAAATTTTCCAGCAAATAAAATCCAAAGTCGATGCCCGCGTTAACTTAAC GCAGTTGCTCGAATGCCAAACAACGTGGGATTTGCTTCTCTATTTGAAATCGCAAAGAGACATACCGCAAATCGAGCCAGAGCCGACACATCAGCAATTAGCTGATTTCT CTTTGTCAAATGACATTCTTTTCCCAGAGCTAATCCAGCTTAATTCGAGTACCAAGGGAAGACCTGTATTCTGGTTTCATGGGATCGCGGGAGTTATGGTTTACGCGCCCG TTGCGGAAAAAAGCCAGCGCCCGTTCTATGGCATTCAGCCCTGGAGTTGGATCAACAAAACCGCACTCCCTTCTCATATTCAACCTATGGTGAAACGTTATCTTGATGCCA TACGATCCGTACAGCCGGAGGGACCCTACGATTTTGGTGGCTATTCGCTGGGTGGGGCGTTAGCTTATGAAGCAACCCGTCAATTACAGGAACTGGGAGAGCGCGTTAAC ACCATTGTGATGGTGGATACGATACAGAATAACCCTGGTGCCAAGAATAAATACTCGCGCAAAACAGACTATCTGGTTACCTTGAATCGAGTGTTGGCCTTATCTGCATG GCAGCGTTCAGAAAAAACTTTGGAGGAAATGCTGATTCGTTCAGATGAACTGGATGCAAGCCTAAGTGACGACAAATATTTAAGTCAATTGATCGCCTTAGCTCAAAAAC GAGGATTAGTGAAAGCAGAGACAGAAATACGTGCCGACCTCGAATATACAACTCTTGATGAATTTTATCAAGCCGATCCCTTTACTATTATGCCTCTGCCCGATCCA AACTTGGTAAATTGCTATTATTTCCGCAACAAGAACGGATCAATGTTGGGTGATCAGGCTCCCTATTATTTTGCGATACCGCAGGATAGAGAAAAATTTGCTACCGATCAG GCAGCCGTTTCGCGGCAATGGGAAAACAACTTGCCGAACCTGCAAATTATTGATATTAATTCGTCCAGTCACATGACCATGTTCTCAGAAGAGAAGTCACGCGATACGAT TCTTGATTTTTGTAAAACCCTGTATTCTGAAAATGGACTTTTGACCTAGGTTTAAACAT
pUC57_*Nos*_ACP_ACP_TE	ATGGCTAGCAATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCA GAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCAC ATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATAC ACTTCAAACCGGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAGGCGATAAG CTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCT GAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATC CATTTGAAAAGAGAAGCTGCTGGAACGGCTTTCAAGCAGTAGAGAATACGCCTTCTGTTTCACAGGATGAGCGTATCAACAACAGCAGCGATCATCACATATTAGCAAAT GTACTAGGGATGGCTCCGGATGAACTGCAGTTTCATAAGCCATTGCAGCAGTATGGATTTGATTCAATTTCTTGCATACAGTTATTACAGCAATTGCAATCAAAGGTGGAC CCTCTCATTGTCTTGACGGAGCTTCAAGCATGCCATACTGTTCAGGACATGATGGACTTGATCGCAAAGTCGTTACAGGATAAAGAGGCACTGAAAAAGATAGCATCTGA AGTGTTCAACGGAGTCCGCTCAGATCATATTTTCCCAAAAATTAGTCAACCTGCTGATTCTGGTCTGTACGAACAAGCGATCGCAACTACAAATATACTTCACAAAGACC ATAAGTTAGGGCTATTACCCGCAGAGGGCATAGATGTTTTTCGTCGTCTTTTGTGTAATACACTGCCTCAAGTTATTGTATCGACACGAGATTTACATTCTCGGATTGAGC AAAAACCAAGTCTCTACACTACTTAGTTTTTTTAGAAGAATCATCGAAGGAGGATCTACCCAAACAGGTACACCAACGACCATCAATAACTCAGGCTTATATTGTTCCGAGA

	AACAAGCTTGAGCAGATTATTGCTGGTATCTGGGAGGATGTTCTTGGGTTTGAGCAGGTGGGGAATCCACGATGATTTCTTTATGTTAGGAGGAGATTCCCTTGTAGCAGTG AGGGTCATGGCTCAAATAAAGAACATATTTCAGGTAGAGTTACCCTCGGCTACCCTGCTCCAGAAGCCCACTATTGCCCAACTATCCCAACTGCTCAAAGCCAGTAATCA TTCTACTATTAATAGACATCTCCAGAAATTAGGTTTTCGCCAAGACCAGATAAAAGCTTCAGATTCTTTTTCAAAGATGTCTAATGCGATTAATTCAGTATTAATACCCATT CAGCTTAAAGGAGAACAAGTTCCTCTGTTCTGTGTTCATCCGGTAGGTGGGAACGTACTGTGCTATGCAGCACTCTCCCAACAGTTAGGGCAGGAACATCCTATCTATGG ACTGCAATCATTAGGCTTAATAGAACAGCAGCTGACTTTGAACACCATTGAGGCAATGGCCAGTCACTACCTTGAGGAAATCTTGACGGTGCAACCTCAGGGTCCTTATC ACCTTATAGGCTGGTCGATGGGAGGTATAATTTCTTACGAAATAGCACAACAGCTTATTCATCGAGGAGAGACAGTTGCATTGTTAGTTCTGATTGATAGCCATATCCCGA CTGCTGAAGACATACCAGGCAAGGATCAACTCCTTGTGGAGTTCATAAAAGATTTAGCTGGGCAGTCGAGACAAAGCTTGCACTCAGTTCTAGAGTCACTAGAAGGAAA ACTAGTTTCTTTAGAGCAAGTATTCAAACTTACCCAACATCATCAACTTTTCCCATCTGAGTTGTCTTTGGAGCAATTACAGCGCTTATGGCGGGTTTTTCAAGTAAATATG CTGGCCTTAACCCAGTATCACCCGCAGCCATATAACGGGGCAGTTTTATCACTTGTAGCAAGTGAGTCAGTCGCACAAGCCTCGCATACAGTAAGTAATGGTGGAGACAG AGTTATATGGGACACGCTAATTAAAGGCGAACTGAGTGTACACATCATTAATGGAGATCATTTTACCCTGATGCGAGAGCCAAACCAAGTTGCCGAAATTGCATACTACC TTGGTTGCTATCTCTAGGTTTAAACAT
pUC57_*TtoF*_ACP_PCP_TE	ATGGCTAGCAATGACCAGGTGTCATTAAAGGGCACCCGTTTCCGAGTGACTGCCGAGAATCAGCAATGGAGGGATTTAAGTGACGATGCAGGCAAAAAAATTCCGCGCA GAGCGAGTATCAACAGCTATAGCTTTGGAGGCGTAAATGCGCACGTCATTTTAGAAGAATATATTCCTTTACCAAAACCACCGGTTAGTATGAGTGAGAATGGTGCCCAC ATTGTAGTTCTTTCTGCAAAGAATCAAGACAGGCTAAAAGCAATTGCTCAACAGCAGCTTGACTATGTGAATAAACAACAAGAACTGTCATTACAAGATTATGCTTATAC ACTTCAAACCGGCCGAGAGGAAATGGAAGACCGCCTGGCGCTCGTCGTCCGCAGTAAAGAAGAACTGGTAATCGGCTTGCAAGCCTGCTTAGCAGAAAAAAGGCGATAAG CTGAAGAGTTCTGTACCTGTCTTTAGCGGAAATGCAGAAATGGCTCGTCAGATCTCGAAGCCTTGCTGGATGGTCCATTAAGAGAAATGGTGATCGAGACTTTGTTGTCT GAAAACAACCTTGAAAAGATCGCGTTTTGCTGGACAAAAGGGGTGCAAATCCCATGGGAAAAGCTTTATCAAGGAAAAGGTGCCCGCAGAATACCGTTGCCAACCTATC CATTTGAAAAGAGAAGCTGCTGGAACGGCTTTCAAGCAGTAGAGAATACGCCTTCTGTTTCACAGGATGAGCGTATCAACAACAGCAGCGATCATCACATATTAGCAAAT GTACTAGGGATGGCTCCGGATGAACTGCAGTTTCATAAGCCATTGCAGCAGTATGGATTTGATTCAATTTCTTGCATACAGTTATTACAGCAATTGCAATCAAAGGTGGAC CCTCTCATTGTCTTGACGGAGCTTCAAGCATGCCATACTGTTCAGGACATGATGGACTTGATCGCAAAGGAGGCTAAAGTTGAGGTTCAAGTTCCTTCTCAAGTTCAGACC GAGGTTCAGACAGAGCCCCTTTCTCAAAGCGTTGGCGCTCAAAGCCCCTCTGGCAAAAACATAGAGGTAGCGCTTACGGATATTATGCTTCTAACCTTAGGGTTGAGGGC AGAAGAAATCAATCTCAATACCCCACTGGTTCAATATGGGGTTGATTCCATCATGTTTGTGCAAATTTTCCAGCAAATAAAATCCAAGATTGATGATCGCGTTAACTTGGG ACAGCTGTTGGAATGCCAAACAATGCAGGATATGGTTCTCTATTTGAAATCGCAAAAAGATACGCCGCAAATTGAACCAGAGCAAGCGGTAAGTTTCTCTTCGTCAAAAG ACGCTCCTTTCCCAGAGCTAATCCAGCTTAATGGGAGTACCAAGGGTAGACCTGTTTTCTGGTTTCACGGGATCGCTGGCGTTAAGGTTTACGAACCCGTGGCTGAAAAA AGTCAACGACCTTTTTATGGTATCCAGCCTTGGAGTTGGATTGACAAAACCCAAATTACTTCTCACATACAACCTATGGTGACGCGTTATCTTGAAGCCATACGTTCCGTG CAGCCAGAGGGTCCCTATGATTTTGGTGGGTATTCCTTGGGTGGAATGTTTGCTTATGAAGCAACTCGTCAATTACAGGAAATCGGAGAGAGGGTTGAGAGCATTGTGAT GGTAGATACATTACAATACAAACCCGGTGAAATGGAGCAATACTCGCGCAAAACAAACTATTTGATTGCCCTGAACCGAGCCTTAACCTTATCTGCATGGCAGGCATCAG AAAAAAACCGTGCAGGAAATGCTGATTAATTCAGATGAACTAGATGCAAACCTCAACAACGAAGAGTATTTAACTCAACTGATTGCCTTGGCAAAACAACGAGGGTTGGT AAAAACAGAGACAGAAATACGTACCAGCCTCGAACAAGCAACAAACATGGATGAATTTTACCAAACCGATCCCTTTACTTTTATGTCGCTACCCGATGCCAACGCGGTCA ACTGCTATTATTTCCGCAATAAGAACGGATCGATGTTGGGCGATCAAACTTCCTATTACTTTGCAACACCAGAGGACAGAGAAAAATTTGCCACAACCGATTGGTCAGCT GCTTCGCAACAGTGGAAAAAGAACTTGCCGAATATGCAAATTATTGATGTTGATTCGTCCAGTCACATGACCATGTTCTCGGAAGAGAAGTCACGCCAGACGATTATTGA ATTTTGTGAAACCCTGTATTCAGAAAAAAGGACTTTCCAATACATTTCTTGAGTTATTTCTTAAAAAAAACTCGGAAAATACATGGTAGCTTGGAACTGGATTTGTAAGTTTA AACAT

Table S5. Restriction cloning performed in this study.

New construct	Backbone	Insert(s)
pMF37_*basC*	pMF37/pDGICZ	pUC57_*basC*
pFusA_*difL*_TE	pFusA/S/C_KanR	pUC57_*difL*_TE
pFusA_*difL*_ACP_TE	pFusA/S/C_KanR	pUC57_*difL*_ACP_TE
pFusA_*difL*_ACP_ACP_TE	pFusA/S/C_KanR	pUC57_*difL*_ACP_ACP_TE
pFusA_*difL*_ACP_ACP_ACP_TE	pFusA/S/C_KanR	pUC57_*difL*_ACP_ACP_ACP_TE
pFusA_*difL*_link_ACP_ACP_TE	pFusA/S/C_KanR	pUC57_*difL*_link_ACP_ACP_TE
pFusA_*misF*_TE	pFusA/S/C_KanR	pUC57_*misF*_TE
pFusA_*misF*_ACP_TE	pFusA/S/C_KanR	pUC57_*misF*_ACP_TE
pFusA_*misF*_ACP_ACP_TE	pFusA/S/C_KanR	pUC57_*misF*_ACP_ACP_TE
pFusA_*misF*_link_ACP_TE	pFusA/S/C_KanR	pUC57_*misF*_link_ACP_TE
pFusA_*misF*_KS^0_ACP_TE	pFusA/S/C_KanR	pUC57_*misF*_KS^0_ACP_TE
pFusA_*pksS*_KO	pFusA/S/C_KanR	pUC57_*pksS*_KO
pFusA_*lumE*_ACP_ACP_TE	pFusA/S/C_KanR	pUC57_*lumE*_ACP_ACP_TE
pFusA_*Nos*_ACP_ACP_TE	pFusA/S/C_KanR	pUC57_*Nos*_ACP_ACP_TE
pFusA_*TtoF*_ACP_PCP_TE	pFusA/S/C_KanR	pUC57_*TtoF*_ACP_PCP_TE

Table S6. Screening primers used for colony PCRs.

Screening primer	Sequence	Primer binding
CP Cat2.2 Fw	CGTGGCCAATATGGACAACTTC	Inside Cm^R marker
CP Cm Fw	GCTTCCATGTCGGCAGAATGC	Inside Cm^R marker
CP2_Spc Fw	AAGTGGGAAGGACTATATTCAAAGG	Inside SpcR marker
CP all Bs Rv	CATCCCGATGGACAAACTTGG	Downstream of DGH
pFusA Sequencing Rv	GTCCTGGCTGGTCTAGACGTC	For sequencing of constructs
PksXTE_qPCR2 Fw	GTGCGCAGGCTCAGCAAATG	mRNA check
PksXTE_qPCR Rv	GAGGAATTTGCCGCTCCCATTC	mRNA check
basC_qPCR Fw	GCTGTAGATGGCTGGTCTGC	mRNA check
basC_qPCR Rv	CTCAAGACGAAACGATACCGTC	mRNA check
difL_TE_qPCR Fw	GATGCTGCAGACGGTGAACATG	mRNA check
difL_TE_qPCR Rv	CATGTCACCCCGTTCGGATC	mRNA check
misF_TE_qPCR Fw	GAGATCCGTGCCAGCCTTG	mRNA check
misF_TE_qPCR Rv	CGTCATGTGATTGGACGAATCG	mRNA check

Table S7. Overview of the cluster used for recombineering in this study. X: unknown domain.

Cluster	Organism	Proposed offloading mechanism	Last module
Bacillaene	*Bacillus*	Linear	(KS^0) ACP TE
Basiliskamide	*Brevibacillus*	Linear, *in trans*?	C
Difficidin	*Bacillus*	Cyclizing	(KS) ACP ACP TE
Misakinolide	*Entotheonella*	Dimerization	(KS^0) ACP TE
Tolytoxin	Cyanobacterium	Cyclizing (lactonization)	(KS^0) ACP TE
Psymberin	Cyanobacterium	Aromatic cyclization	(KS) X DH ACP TE
Luminaolide	Cyanobacterium	Dimerization	(KS^0) ACP TE
Nosperin	Cyanobacterium	Linear	(KS) AT KR ER ACP TE

Table S8. DHB levels normalized to bacillibactin levels as area under peak of extracted ion chromatograms from UHPLC-HRMS analysis, % cluster output (% CO) and fold increases computed for the different mutants.

Strain	DHB normalized	%CO	Fold change
DifL_TE	9.32E+06	10.8 ± 4.2%	2.24
DifL_ACP_TE	7.75E+06	13.2 ± 6.3%	2.74
DifL_ACP2_TE	1.13E+07	14.6 ± 2.8%	3.03
DifL_linker_ACP2_TE	1.44E+05	0.3 ± 0.5%	0.07
DifL_ACP3_TE	7.49E+06	12.9 ± 4.4%	2.67
BasC_in_TEKO	3.64E+06	9.6 ± 3.1%	2.00
MisF_TE	5.42E+05	0.6 ± 0.3%	0.12
MisF_ACP_TE	6.39E+05	0.7 ± 0.5%	0.15
MisF_link_ACP_TE	1.52E+05	0.4 ± 0.3%	0.08
MisF_ACP2_TE	6.73E+05	1.5 ± 1.1%	0.32
MisF_KS0	2.06E+04	0.1 ± 0.2%	0.03
BasC_in_WT	4.53E+07	76.7 ± 21%	15.91
PsyD_TE	6.97E+04	0.8 ± 0.4%	0.16
PsyD_ACP_TE	3.25E+04	0.4 ± 0.4%	0.08
PsyD_ACP_ACP_TE	9.65E+04	0.6 ± 0.4%	0.12
Nos_ACP_ACP_TE	2.27E+05	1.9 ± 1.4%	0.40
Lum_ACP_ACP_TE	2.90E+05	2.9 ± 1.2%	0.61
TtoF_TE	9.09E+05	17.5 ± 8.2%	3.62
TtoF_PCP_TE	6.96E+05	15.9 ± 3.7%	3.29
TtoF_ACP_ACP_TE	2.18E+06	15.6 ± 1.6%	3.24
Δ*pks*	1.57E+02	0.1 ± 0.4%	0.02
Δ*pksS*	2.43E+07	53.0 ± 14%	11.00
ΔTE	1.38E+06	4.8 ± 2.2%	1.00

Supplementary Figures
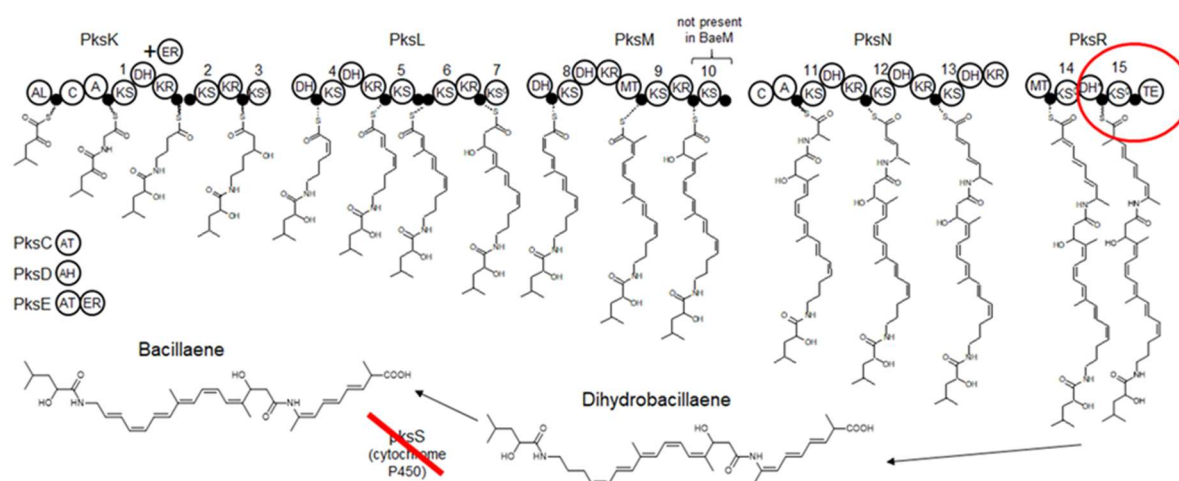
Figure S1. Biosynthetic model of the bacillaene *trans*-AT PKS in *B. subtilis*. The terminal region is marked with a red circle. The post-PKS cytochrome P450 PksS-encoding gene was deleted to facilitate studies on offloading capacity between different mutants. The protein names are indicated above the respective portion of the cluster. AL: acyl ligase, C: condensation domain, A: adenylation domain, KS: ketosynthase, DH: dehydratase, ER: enoyl reductase, KR: ketoreductase, KS^0: non-elongating KS, MT: methyltransferase, TE: thioesterase, AH: acyl hydrolase, black circles: acyl or peptidyl carrier proteins.
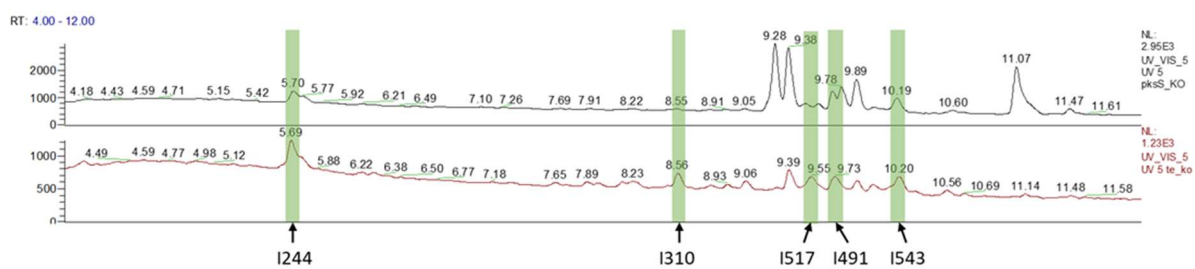
Figure S2. UV traces of WT Δ*pksS* (top) and ΔTE (bottom) culture extracts. The bacillaene peaks can be observed at 11.07 min. Different intermediates are shaded in green and labelled as I+*m/z*.
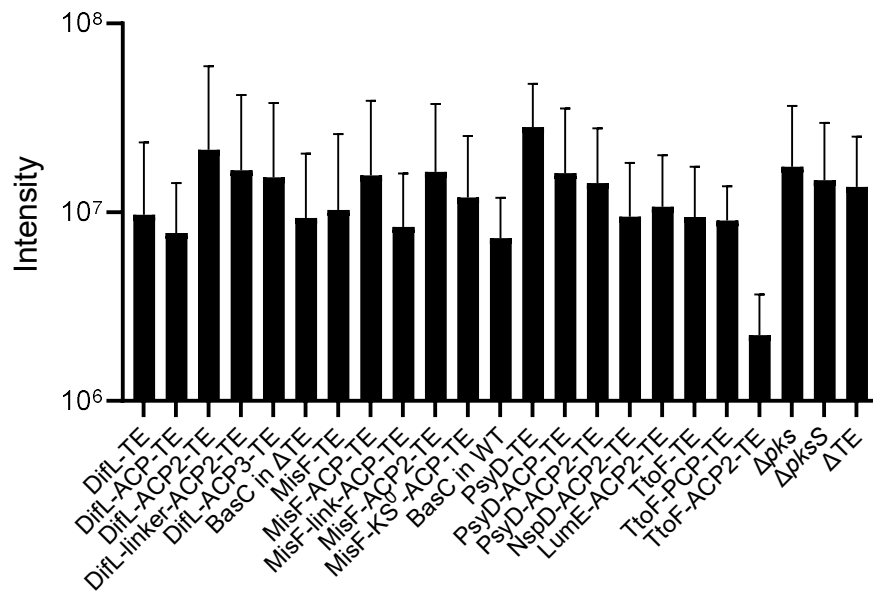
Figure S3. Overview of bacillibactin levels for all measured mutants.
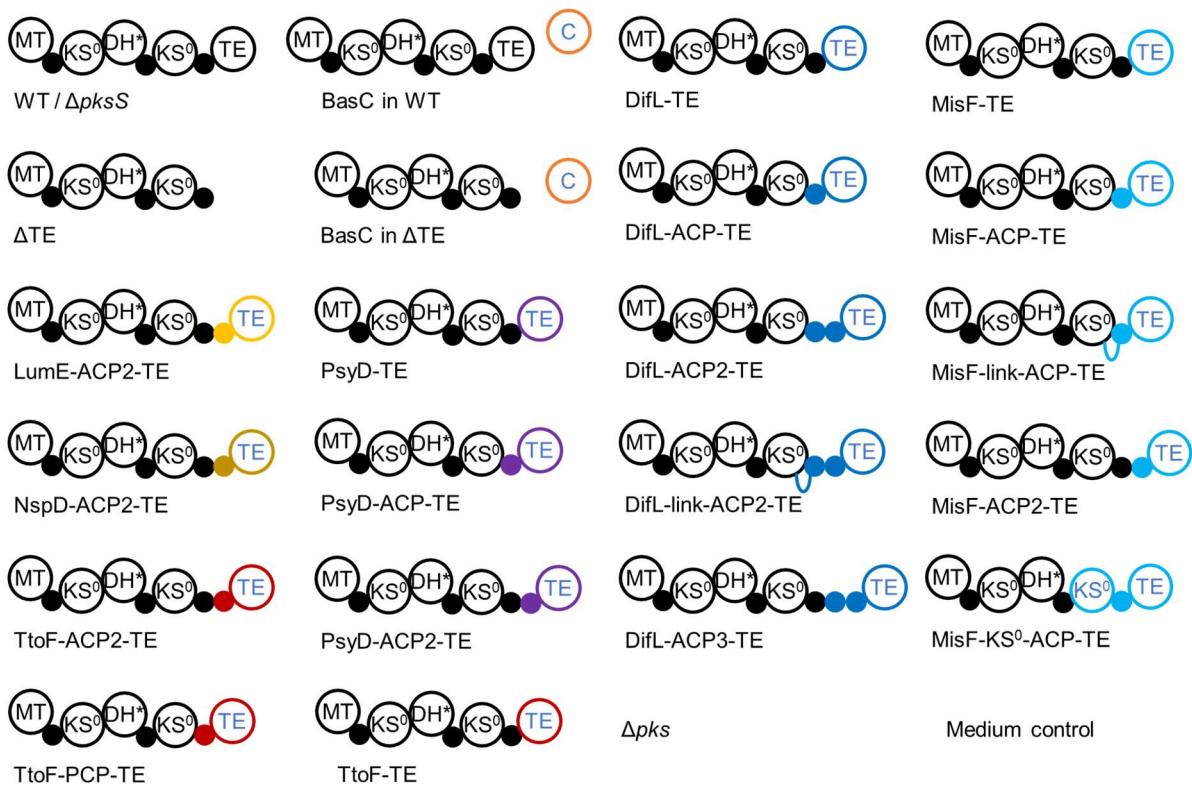
Figure S4. Overview of all the mutants constructed in *B. subtilis*. Colored domains are non-native, half circle represents a linker. Δ*pksS* represents *B. subtilis* lacking the cytochrome P450 responsible for the conversion of dihydrobacillaene to bacillaene.
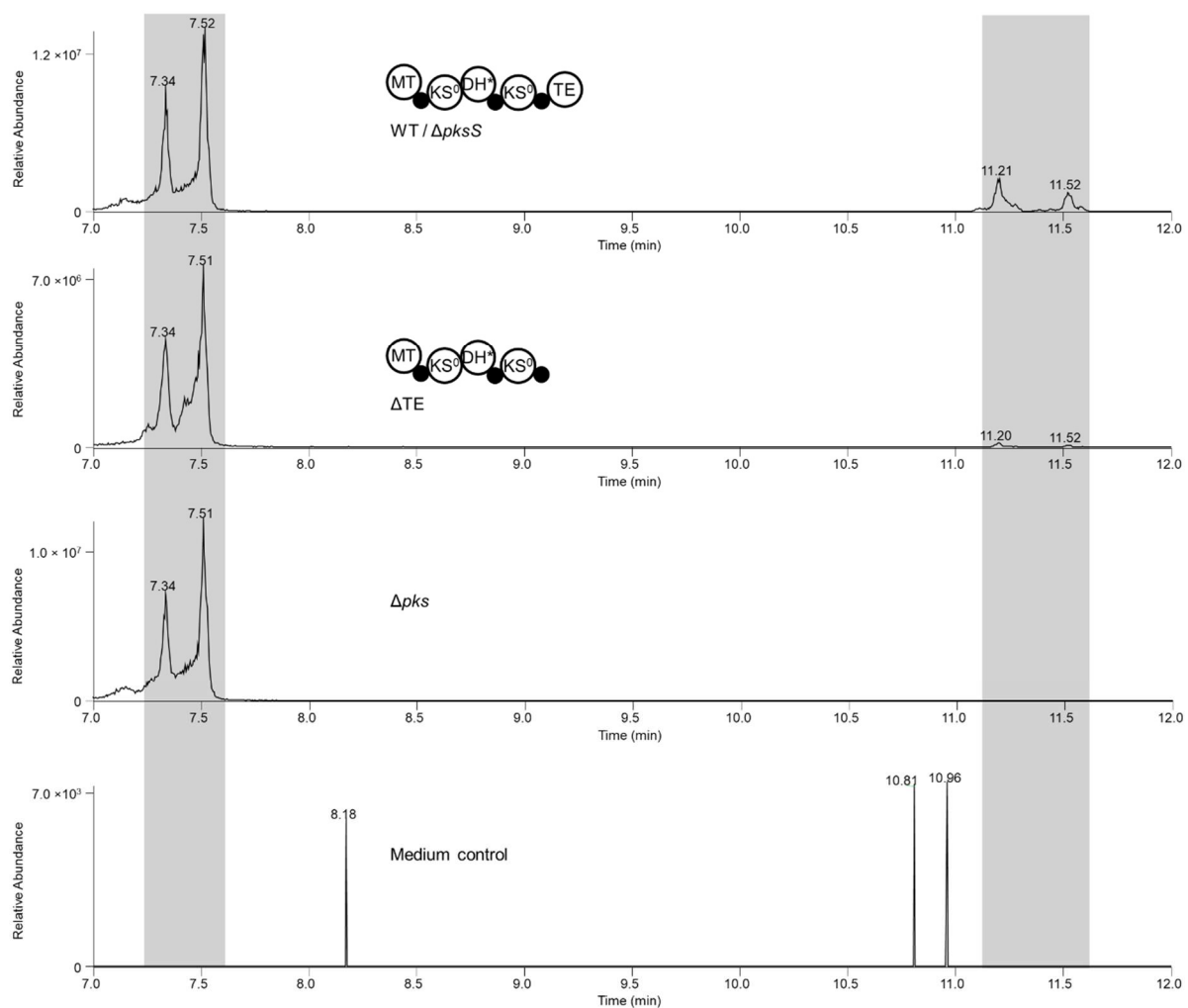
Figure S5. Representative extracted ion chromatograms (EICs) of the WT ($\Delta pksS$), ΔTE, Δpks, and medium control. The peaks at 7.3 min and 7.5 min correspond to bacillibactin (calculated for $[M+H]^+$ 883.2650) used as an internal standard. Dihydrobacillaene (calculated for $[M+H]^+$ 583.3742) is found at 11.2 and 11.5 min. Grey bars mark the regions where bacillibactin and dihydrobacillaene peaks are expected.
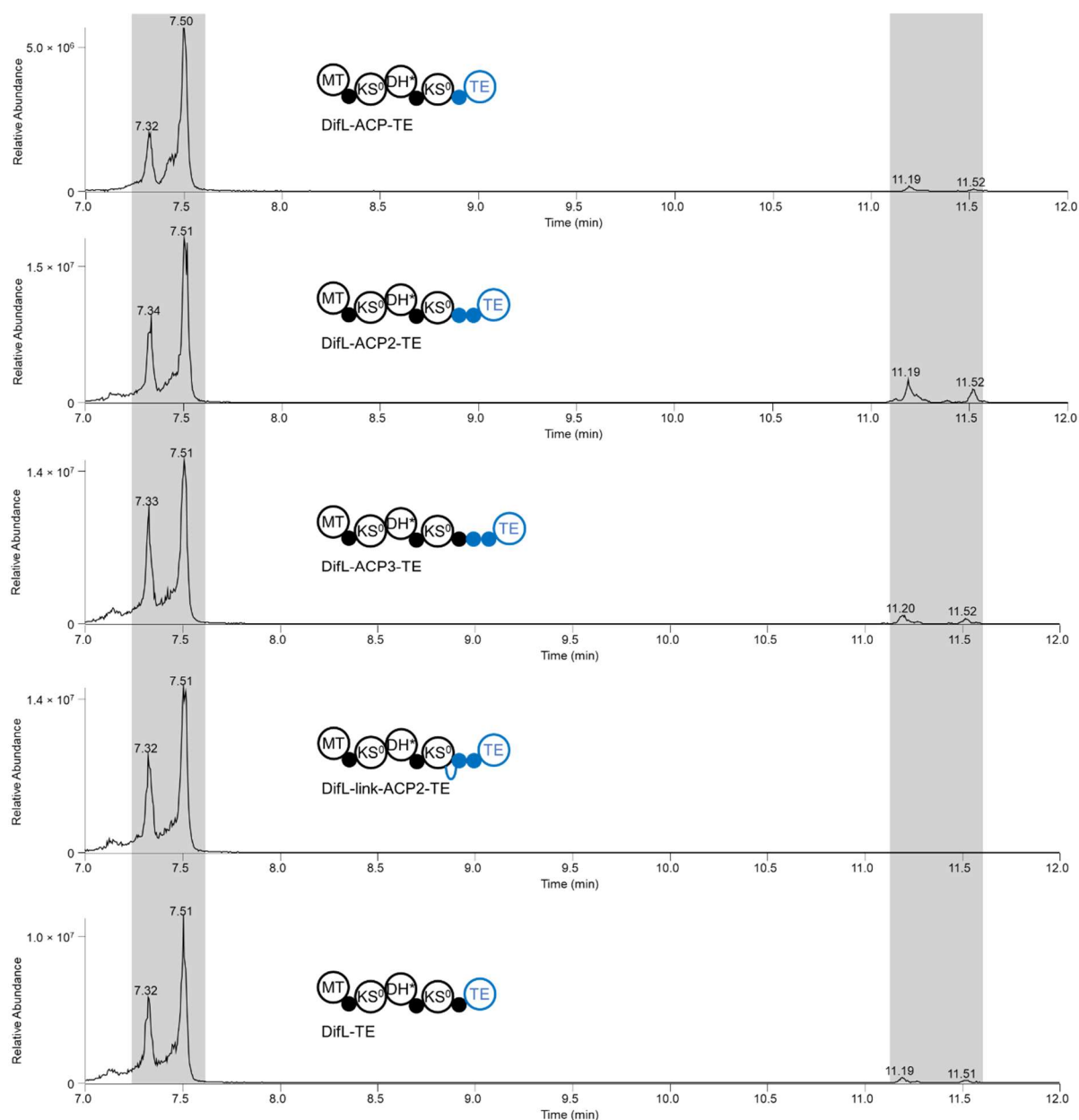
Figure S6. Representative extracted ion chromatograms (EICs) of the difficidin TE swaps. The peaks at 7.3 min and 7.5 min correspond to bacillibactin (calculated for $[M+H]^+$ 883.2650) used as an internal standard. Dihydrobacillaene (calculated for $[M+H]^+$ 583.3742) is found at 11.2 and 11.5 min. Grey bars mark the regions where bacillibactin and dihydrobacillaene peaks are expected.
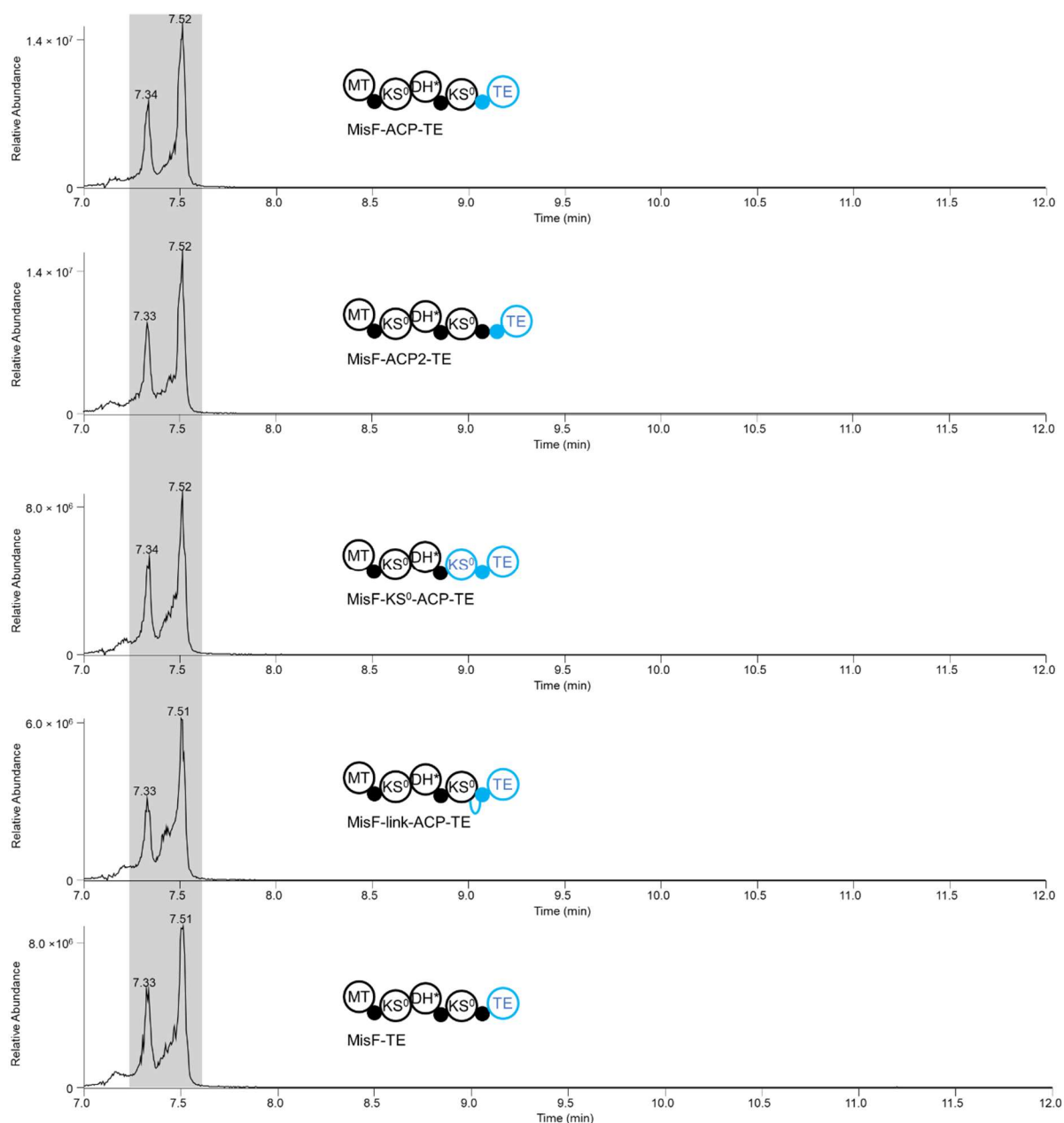
Figure S7. Representative extracted ion chromatograms (EICs) of the misakinolide TE swaps. The peaks at 7.3 min and 7.5 min correspond to bacillibactin (calculated for $[M+H]^+$ 883.2650) used as an internal standard. Dihydrobacillaene (calculated for $[M+H]^+$ 583.3742) would be found at 11.2 and 11.5 min. Grey bar marks the regions where bacillibactin peaks are expected.
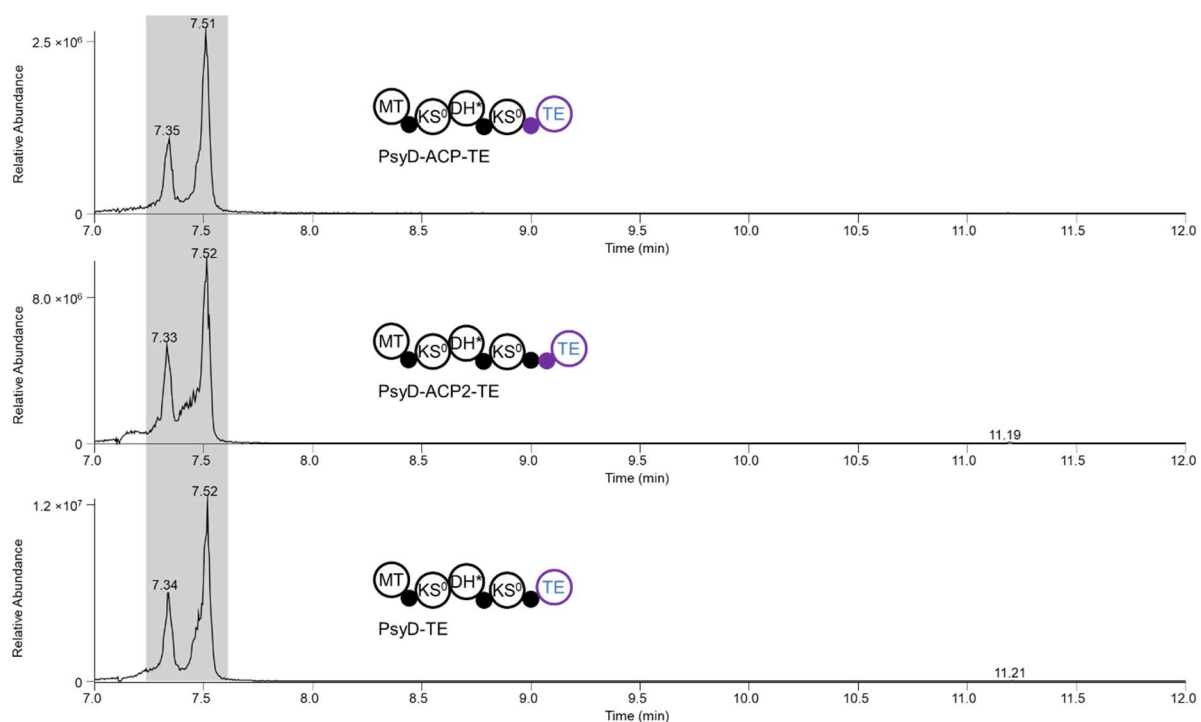
Figure S8. Representative extracted ion chromatograms (EICs) of the psymberin TE swaps. The peaks at 7.3 min and 7.5 min correspond to bacillibactin (calculated for [M+H]$^+$ 883.2650) used as an internal standard. Dihydrobacillaene (calculated for [M+H]$^+$ 583.3742) would be found at 11.2 and 11.5 min. Grey bar marks the regions where bacillibactin peaks are expected.
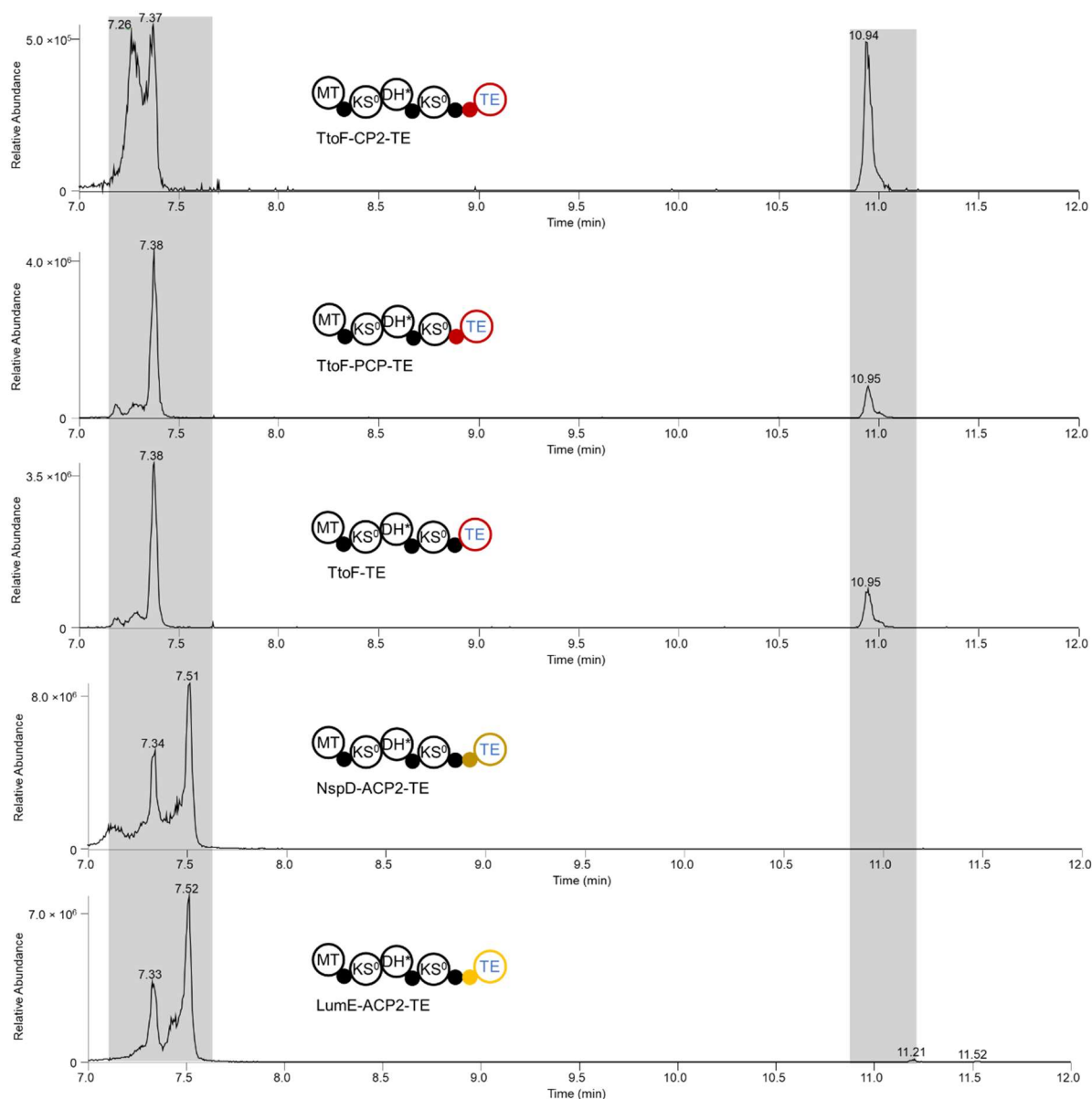
Figure S9. Representative extracted ion chromatograms (EICs) of the tolytoxin, nosperin, and luminaolide TE swaps. The peaks at 7.3 min and 7.5 min correspond to bacillibactin (calculated for $[M+H]^+$ 883.2650) used as an internal standard. Dihydrobacillaene (calculated for $[M+H]^+$ 583.3742) is found at 10.9 min. Grey bars mark the regions where bacillibactin and dihydrobacillaene peaks are expected.
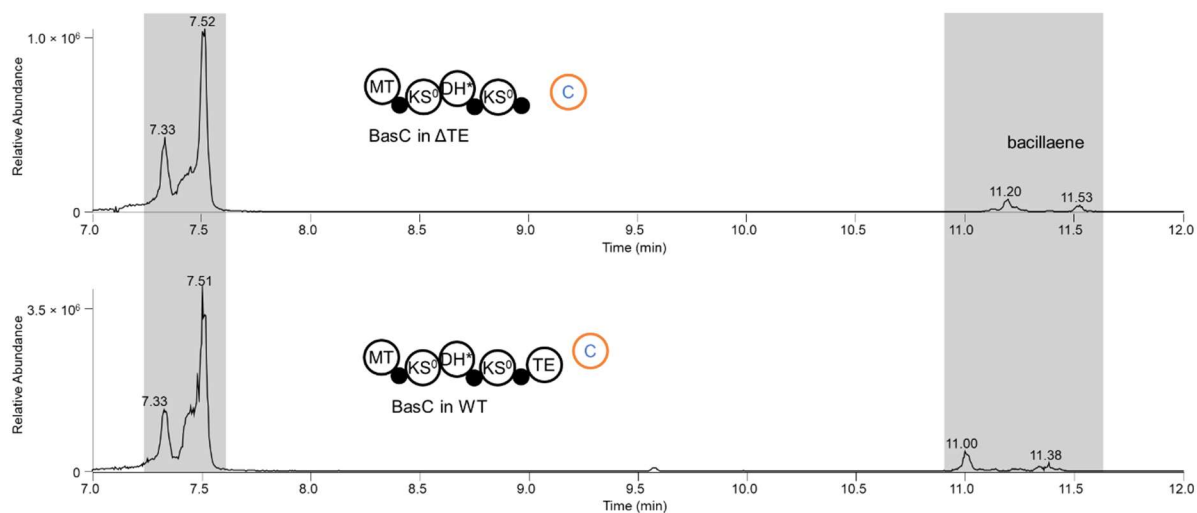
Figure S10. Representative extracted ion chromatograms (EIC) of the TE swaps with added BasC. The peaks at 7.3 min and 7.5 min correspond to bacillibactin (calculated for $[M+H]^+$ 883.2650) used as an internal standard. Dihydrobacillaene (calculated for $[M+H]^+$ 583.3742) is found at 11.2 and 11.5 min. Bacillaene (calculated for $[M+H]^+$ 581.3585) in WT is found at 11.0 and 11.4 min. Grey bars mark the regions where bacillibactin and dihydrobacillaene peaks are expected.
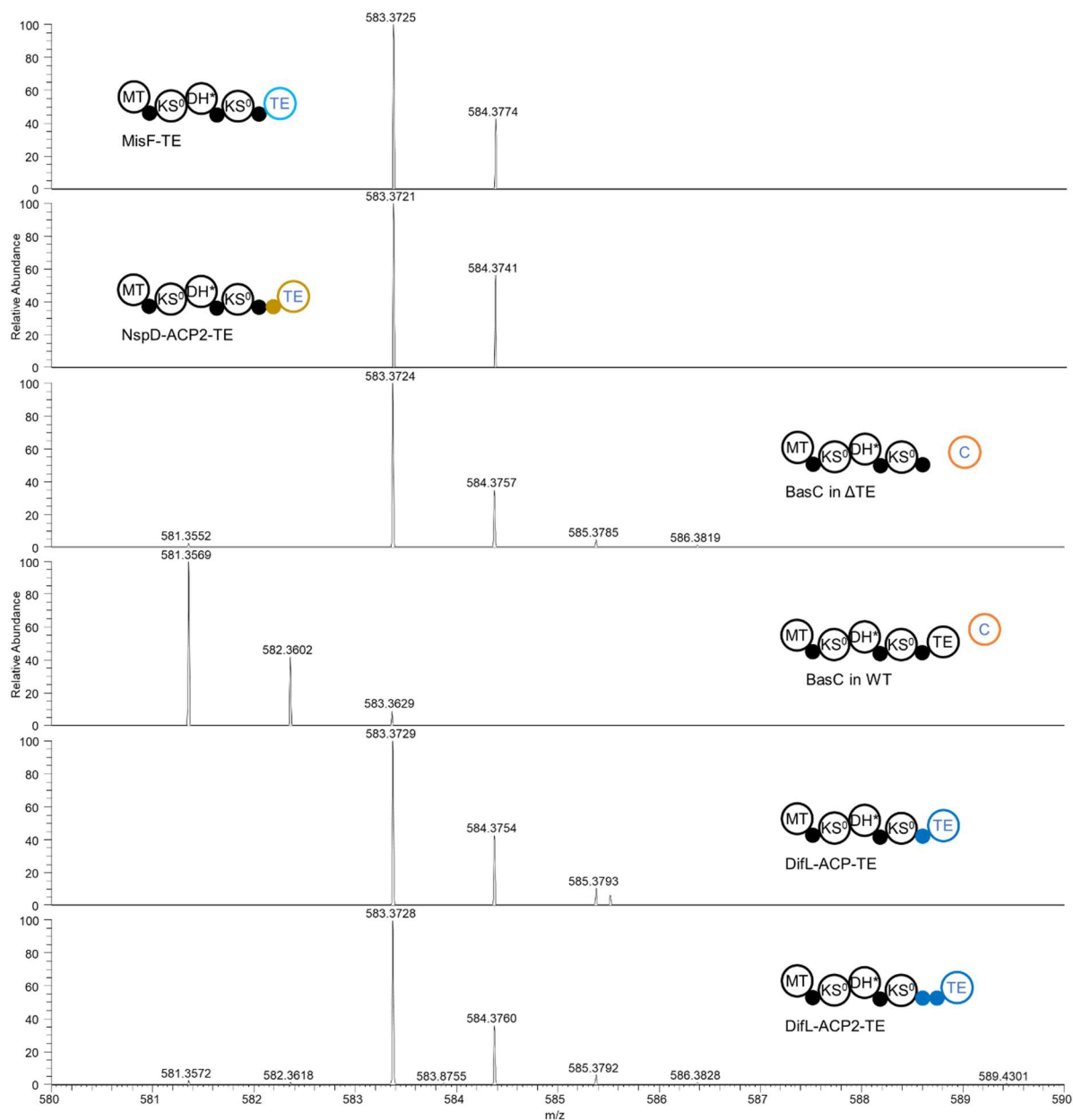
Figure S11. Representative mass spectra corresponding to the dihydrobacillaene (calculated for [M+H]$^+$ 583.3742) peak for the different mutants.
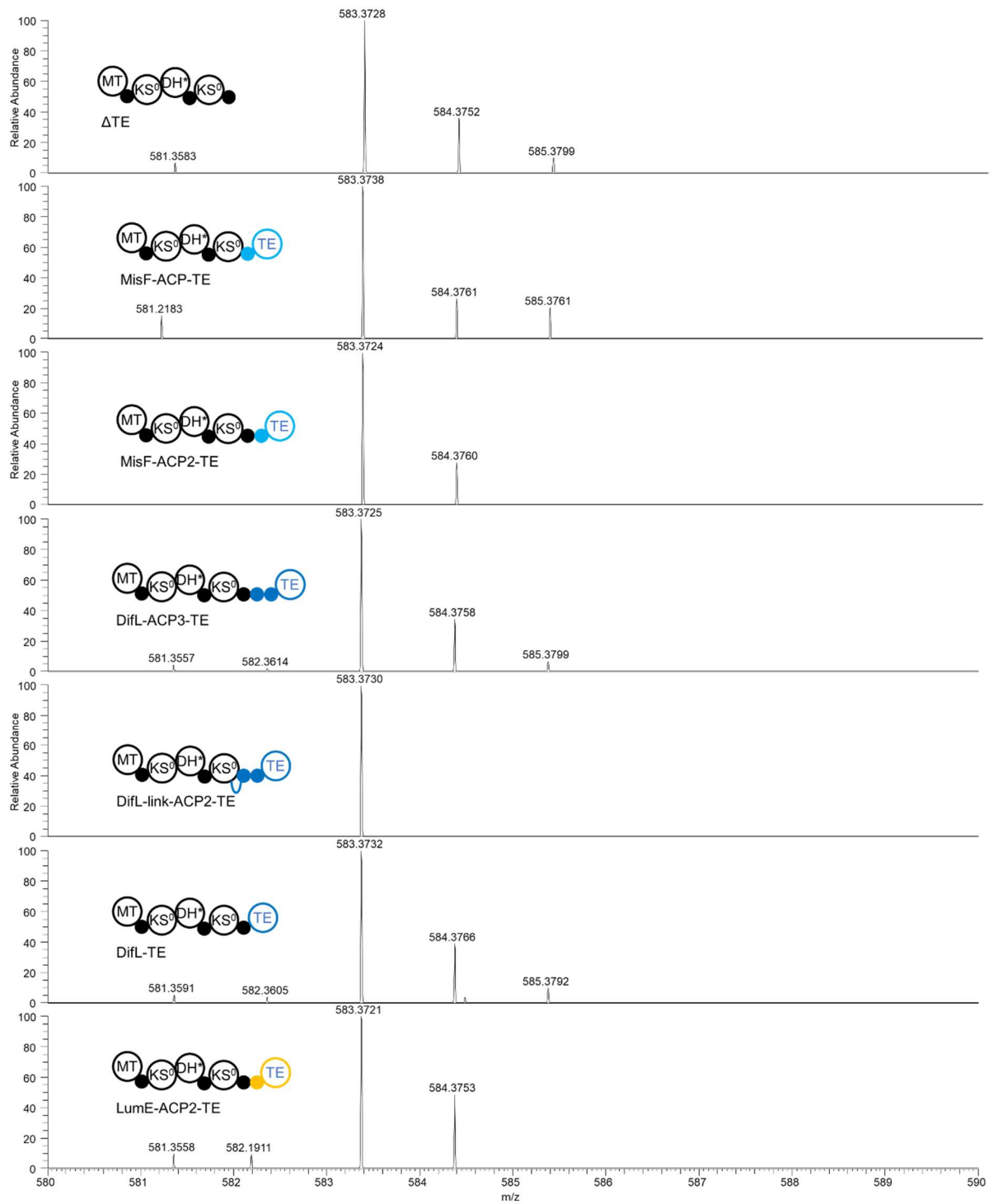
Figure S11 continued.
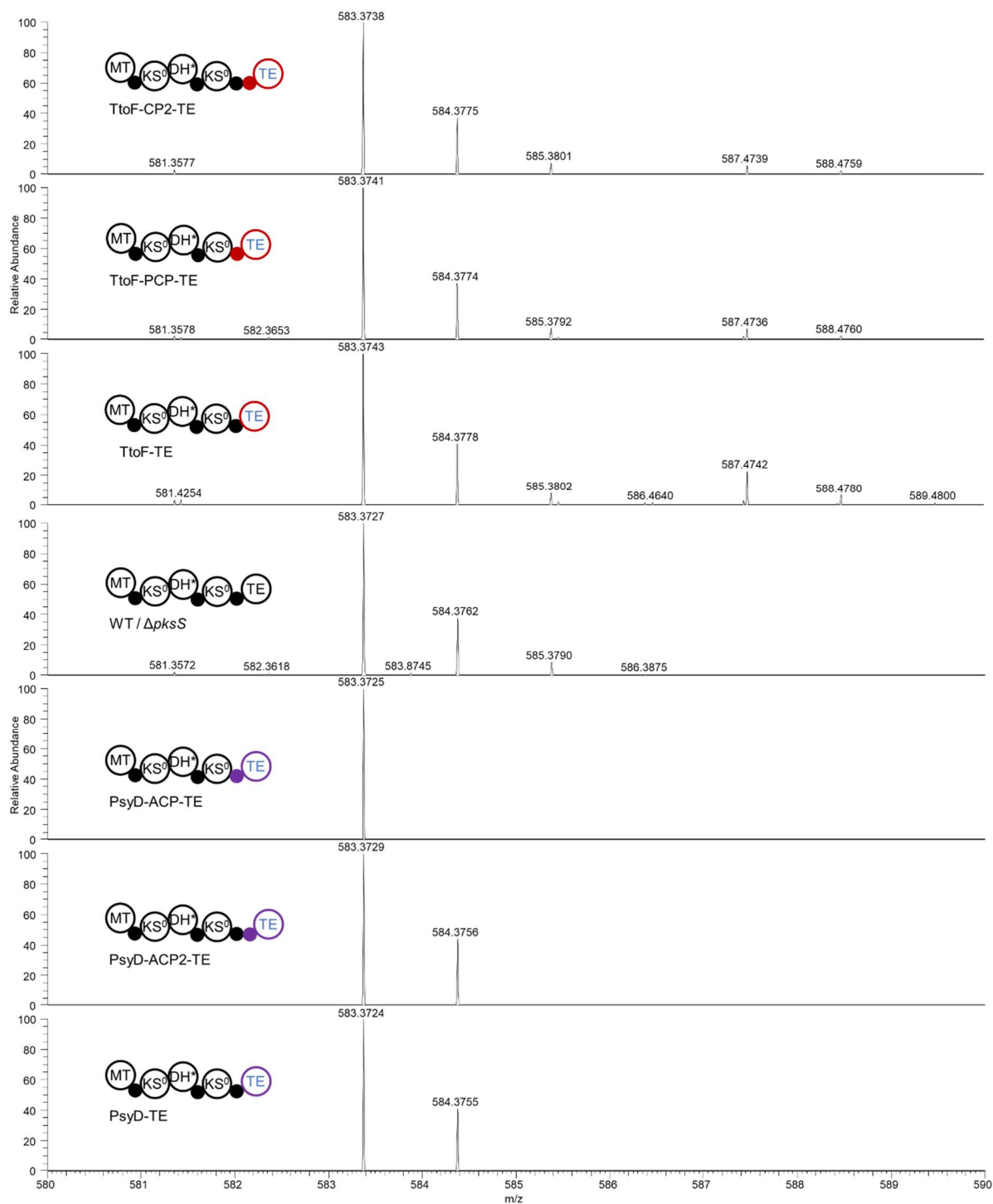
Figure S11 continued.
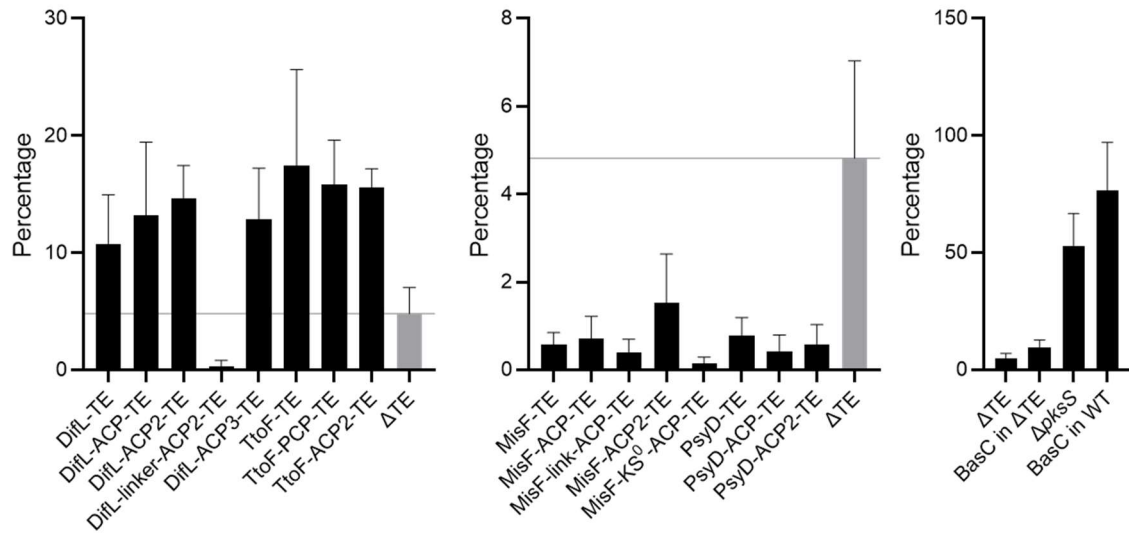
Figure S12. Percentage total cluster output (% CO) computations for the additional mutants. ΔTE is included as a reference and marked in grey.
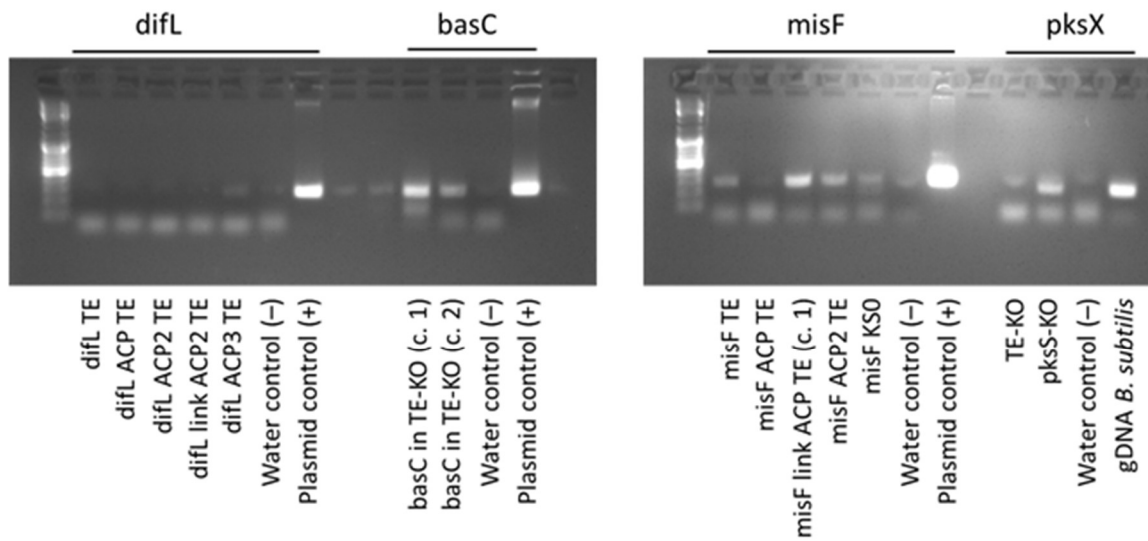
Expected size of TE test fragments: 250-300 bp.
GeneRuler DNA Ladder Mix

Figure S13. Agarose gel of PCRs for mRNA expression control of selected terminal mutants.
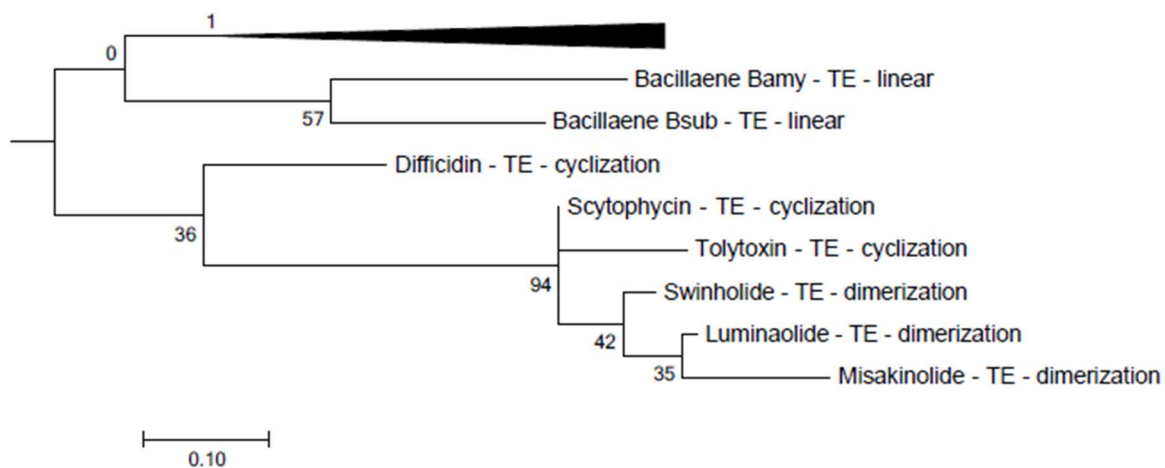
Figure S14. Excerpt of a maximum likelihood phylogenetic tree of TEs and terminal C domains from 79 *trans*-AT PKSs. The group of PKSs responsible for the biosynthesis of actin-binding natural products form a distinct clade with tolytoxin as its most basal branch. The sequences are annotated as compound name - type of offloading enzyme (TE or C) - mode of release according to the corresponding biosynthetic model.

References

[1]   M. A. Konkol, K. M. Blair, D. B. Kearns, *J. Bacteriol.* **2013**, *195*, 4085-4093.
[2]   R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straightt, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506-1509.
[3]   a) C. Engler, S. Marillonnet, in *cDNA libraries*, Springer, **2011**, pp. 167-181; b) T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, D. F. Voytas, *Nucleic Acids Res.* **2011**, *39*, e82-e82.
[4]   R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792-1797.
[5]   S. Kumar, G. Stecher, K. Tamura, *Mol. Biol. Evol.* **2016**, *33*, 1870-1874.
[6]   G. Zappia, P. Menendez, G. D. Monache, D. Misiti, L. Nevola, B. Botta, *Mini-Rev. Med. Chem.* **2007**, *7*, 389-409.
[7]   A. Loeschcke, A. Markert, S. Wilhelm, A. Wirtz, F. Rosenau, K.-E. Jaeger, T. Drepper, *ACS Synth. Biol.* **2013**, *2*, 22-33.
[8]   M. Steinmetz, R. Richter, *J. Bacteriol.* **1994**, *176*, 1761-1763.
[9]   X. Yan, H.-J. Yu, Q. Hong, S.-P. Li, *Appl. Environ. Microbiol.* **2008**, *74*, 5556-5562.

Chapter IV

Manuscript in preparation

Evolution-Guided Engineering of *trans*-Acyltransferase Polyketide Synthases

Hannah A. Minas[1], Mathijs F.J. Mabesoone[1], Roy A. Meoded[1], Sarah Wolf[1], Stefan Leopold-Messer[1], Silke I. Probst[1], Silke Reiter[2], Nancy Magnus,[3] Birgit Piechulla,[3] Allison S. Walker[4], Eric J.N. Helfrich[5,6], Jon Clardy[4], and Jörn Piel[1*]

[1]Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zürich, Vladimir-Prelog-Weg 4, 8093 Zürich (Switzerland).
E-mail: jpiel@ethz.ch

[2]Bioprocess Laboratory, Eidgenössische Technische Hochschule (ETH) Zürich, Mattenstrasse 26, 4058 Basel (Switzerland).

[3]Institute for Biological Sciences, University of Rostock, Albert-Einstein-Straße 3, 18059 Rostock (Germany)

[4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, United States.

[5]Institute for Molecular Bio Science, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

[6]LOEWE Center for Translational Biodiversity Genomics (TBG), 60325 Frankfurt am Main, Germany

Author contributions

HAM, RAM, SR, and JP designed the research. HAM, MFJM, RAM, and SW performed cloning, expression, HPLC-MS analysis and analyzed the data. HAM, MFJM, SLM, and SIP performed compound purification and NMR experiments. MFJM and ASW performed statistical analysis. HAM and SR performed cloning, expression and HPLC-MS analysis of *Bacillus* constructs. NM and BP constructed the *Serratia* deletion mutants. HAM, MFJM, and JP wrote the manuscript and the supporting information with input from all authors.

Part of the *Bacillus* constructs used in this chapter were already presented in the thesis of Silke Reiter.

Abstract

Multimodular polyketide synthases (PKSs) are biosynthetic megaenzymes that produce a wide range of natural products. Many of these compounds and their improved variations are in clinical use as antibiotics, antitumor, immunosuppressant, or other drugs but chemical synthesis of the complex scaffold structures is often laborious and uneconomic. The high chemical diversity of complex polyketides and their production by modular assembly lines has raised a longstanding interest in PKS engineering for directed bacterial polyketide production. While earlier efforts have successfully produced hybrid assembly lines for the canonical *cis*-acyltransferase (AT) PKS family, modular engineering of the biochemically less studied *trans*-AT PKSs has not been reported. Here we present successful construction of functional hybrids based on the oocydin assembly line of *Serratia plymuthica*. Informed by patterns learnt from *trans*-AT PKS evolution and statistical coupling analysis, we identified sites for non-productive and productive PKS fusions. Experimental validation indicated improved performance of a newly identified fusion site compared to a commonly used *cis*-AT PKS site. These results suggest a broader scope in engineering *trans*-AT PKSs, enzymes of unparalleled biochemical diversity, to access complex molecules. They furthermore highlight statistical coupling analysis as a potential tool to tailor other modular enzyme families.

Introduction

Molecules evolved in bacteria for chemical defense and warfare are among the most valuable foundations for antibiotics and other drugs. However, the discovery of new antibiotics has declined after the "Golden Age of antibiotic discovery".[1] With multi-resistant bacteria on the rise, there is an urgent need to expand our repository of antimicrobial drugs.[2] While traditional discovery strategies remain indispensable, approaches to diversify privileged scaffolds have become increasingly more important.[3] Structurally less complex scaffolds inspired by natural products can be synthesized and varied at large scale for drug development.[1] For more complex leads, semisynthetic approaches and feeding of non-natural building blocks have produced valuable new structures, but these strategies are often expensive or offer only limited expansion of known pharmacophores. To overcome this limitation, engineering of the enzymatic machinery underlying important bacterial natural product families has been a long-standing dream.[4] This strategy has led to some success for a few compound types. However, designing scaffolds for many of the most diverse natural product classes remain challenging.[5]

A rich source of lead compounds are complex polyketides biosynthesized by bacterial multimodular polyketide synthases (PKSs).[6] In these several million Dalton megaenzymes, thioester-bound intermediates are shuttled through a sequentially organized series of domains organized as modules and elongated with small acyl units.[7] In canonical *cis*-acyltransferase (*cis*-AT) PKSs, modules minimally contain a ketosynthase (KS), an AT, and an acyl carrier protein (ACP) domain. The AT loads the ACP with a malonyl-CoA-derived building block that is coupled by the KS to the growing polyketide chain through a Claisen-like condensation. Optionally, a series of reducing domains can further modify the ketone moiety resulting from this condensation. Ketoreductases (KR), dehydratases (DH), and enoyl reductases (ER) can act consecutively to introduce β-hydroxy, α,β-double bond and fully saturated intermediates. The final intermediate is then usually released from the assembly line by a thioesterase (TE) domain. Finally, the product can be further modified by various post-PKS modifications, such as halogenations or cyclizations.[8] The sequence of *cis*-AT PKS modules is usually collinear with the core polyketide structure, a feature permitting structural predictions from the PKS architecture.[9]

The pharmacological importance of polyketide scaffolds along with their modular architecture have long triggered investigations toward engineered PKSs.[4a, 10] The bacterial production of polyketide scaffolds may help develop new natural product-derived compounds in an economical and eco-friendly fashion.[11] Despite great advances in the characterization of PKS function and structures, rational engineering of these modular megaenzymes to access designer metabolites has remained challenging.[4a, 11-12] In several studies, the *cis*-AT deoxyerythronolide B synthase (DEBS) was used as a model system and tailored using *in vitro* and *in vivo* expression.[4a, 10-11] Changes in stereochemistry have been introduced by the exchange of KRs, methylation patterns were altered via the replacement of ATs, and truncated assembly lines producing cyclic triketides were built.[11, 13] Further, this setup was harnessed to successfully recombine modules from different polyketide biosynthetic assembly lines. [4a, 11, 14]

With first efforts in *cis*-AT PKS engineering coming to fruition, recombineering in a second family of PKS has so far remained unexplored. The more recently described *trans*-AT PKSs rely on the same biosynthetic principles but display a vastly expanded functional repertoire.[15] Apart from the eponymous *trans*-acting ATs, additional enzymes can be involved in the transformation of the nascent polyketide. These can either be module-integrated or act *in trans*, which complicates efforts to predict the biosynthetic activity of modules.[15c] Careful analysis of *trans*-AT PKS diversity revealed a hallmark of KSs that helped decipher a prediction strategy as well as potential recombineering principles. While *cis*-AT PKSs seem to evolve in a vertical manner,[16] horizontal gene transfer of DNA encoding PKS sections seems to play the leading role in the evolution of *trans*-AT PKSs.[15b, 16] This extensive

horizontal gene transfer becomes visible when studying the phylogeny of KSs. In *trans*-AT PKSs, KSs do not clade with other KSs from within the same biosynthetic assembly line as is the case for *cis*-AT PKSs.[16] Instead, KSs from different assembly lines clade together that accept intermediates with similar modifications introduced by the upstream domains. One possible explanation for the co-evolution of KS with their upstream domains is to ensure correct domain contacts as the AT closely interacting with KS in *cis*-AT PKSs is not present as a module-integrated domain in *trans*-AT PKSs.[17] Further, exchanges of larger module series have been observed to occur between otherwise unrelated *trans*-AT PKSs.[15b] Such natural hybrid assembly lines are found in clusters from a wide range of organisms.[15b, 15c, 18] Even though the underlying principles of such a mosaic-like evolution are not yet fully understood, their natural occurrence provides a hint that artificial recombineering may be tolerated.

Here we present an approach to produce hybrid *trans*-AT PKS assembly lines following nature's evolutive principles. Despite their more complex appearance, we try to harness the tendency of these megaenzymes to naturally recombine. With the help of bioinformatic and statistical analysis we identify a fusion point for recombination of modules and produce hybrid assembly lines *in vivo*. Apart from the point of recombination, KS compatibility at the sites of fusion turned out to be crucial. Our results provide the first successfully created hybrid *trans*-AT PKSs.

Results

The oocydin PKS in *Serratia plymuthica* as a platform for hybrid production

As a model system for engineering studies, we considered the oocydin biosynthetic pathway in the bacterium *Serratia plymuthica* (*S. plymuthica*) 4Rx13. Oocydins and the related biselides and haterumalides have been isolated from a wide range of sources, including sponges, ascidians, and plant- or nematode-associated bacteria. The *ooc* PKS for the biosynthesis of these differentially bioactive compounds is one of the best-characterized *trans*-AT PKS assembly lines.[19] A number of non-canonical modules install modifications rarely seen in polyketide synthases, such as α-hydroxylation, a vinyl chloride moiety, and *O*-acetylation.[19]

In wild type *S. plymuthica* 4Rx13, extraction of cultures grown in enriched potato broth and subjected to ultra-high performance liquid chromatography/high-resolution mass spectrometry (UHPLC-HRMS) has previously yielded different oocydin congeners (Figure 1A). The final polyketide oocydin C can interconvert to the readily observed oocydin A and oocydin B through enolether or ester hydrolysis, respectively. We have recently shown that the genetic accessibility of *S. plymuthica* 4Rx13 allows for the creation of knock-out strains. These can be complemented with plasmids containing the deleted genes to restore functionality of the knock-out strains. Previously we have shown that deletion of a PKS gene region encoding the terminal offloading domain OocS$_C$ with homology to non-ribosomal peptide synthetase condensation domains leads to the release of intermediates throughout the assembly line.[19] Using UHPLC-HRMS, chlorinated late-stage intermediates can be detected in extraction mixtures based on their characteristic chlorination isotope pattern. In the oocydin *trans*-AT PKS, the chloride is introduced by the *trans*-acting Fe(II)/α-ketoglutarate-dependent halogenase OocP in module 11 together with the auxiliary protein OocQ.[19]
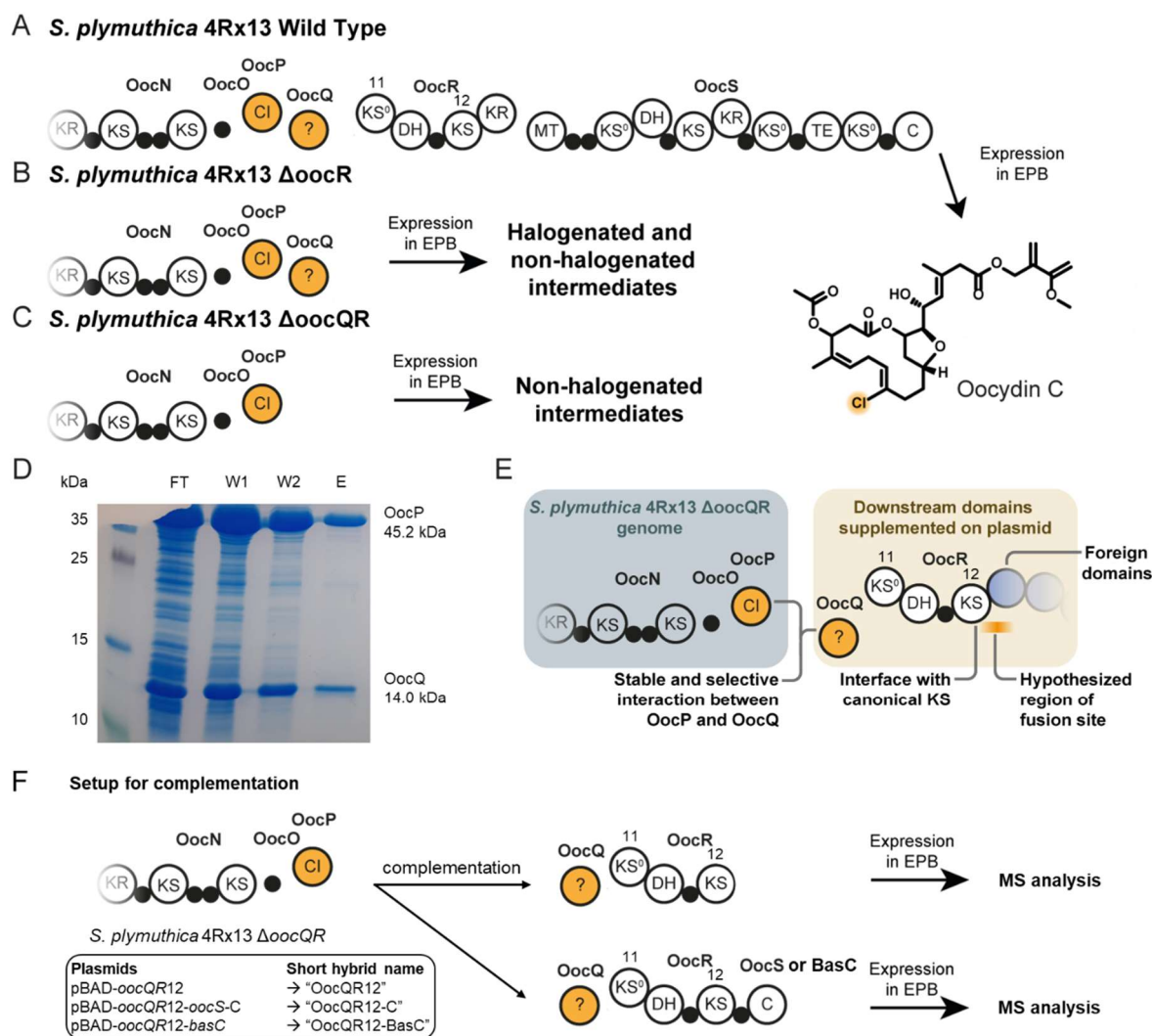
Figure 1. Docking between the halogenase OocP and its putative docking partner OocQ is necessary for the chlorination of oocydin intermediates. A) Domain architecture of the terminal part of the oocydin biosynthetic gene cluster and the polyketide product oocydin C. Small solid circles represent ACPs. C, condensation; Cl, Fe(II)/α-ketoglutarate-dependent halogenase; DH, dehydratase; KR, ketoreductase; KS, ketosynthase; KS^0, non-elongation KS; MT, methyltransferase; TE_B, branching thioesterase. Numbers above KSs indicate their position in the PKS, EPB: enhanced potato broth. B) Cultivation of *S. plymuthica* 4Rx13Δ*oocR* leads to halogenated intermediates. C) Cultivation of *S. plymuthica* 4Rx13Δ*oocQR* does not lead to the formation of halogenated intermediates. D) SDS-PAGE of natively expressed genes for OocP and OocQ. Both proteins are purified as native proteins encoded on one plasmid. FT: flow-through, W: wash, E: elution. E) Strategy for PKS engineering using the interaction between OocP and OocQ as a handle. F) Setup for the first set of control mutants to test the cloning strategy and introduction of the short name nomenclature used in this work.

Inspired by these earlier findings, we constructed *S. plymuthica* 4Rx13Δ*oocR*, a strain lacking the PKS region downstream of the halogenation module (Figure 1B). The product of this truncated assembly line would be smaller. In addition, the incorporation of the chlorine atom would allow for the specific detection of elongated intermediates. After growth of the mutant strain under expression conditions, the culture was extracted and subjected to UHPLC-HRMS analysis. We were able to detect intermediates up to the truncation point, including the first chlorinated intermediate (Figure S2). In line with *in vitro* chlorination assays that used substrate bound to OocO as a surrogate for the biosynthetic thioester-bound intermediate, a strain containing *oocO* and *oocP* but lacking *oocQ* (*S. plymuthica* 4Rx13Δ*oocQR*) was not able to produce chlorinated intermediates (Figures 1C and S2).[19] These studies indicate the need of OocQ for activity by the halogenase OocP. A bioinformatics search of OocQ revealed no described

homologs outside of the oocydin *trans*-AT PKS except for a moderately homologous protein in *Aquimarina* RZ0 (31.58% percentage identity, Table S6). With the production of observable chlorinated intermediates, the modules downstream of OocQ present attractive opportunities to unravel design rules for *trans*-AT PKS engineering in an *in vivo* setting.

Intrigued by the interdependence of OocP and the highly exotic auxiliary protein OocQ, we performed experiments to characterize their interaction. Due to the high number of His in the amino acid sequence of OocP (Figure S3), we first tested whether the native protein can be purified without the addition of a His-tag. To do so, *oocP* and *oocQ* were cloned onto expression plasmids in their native context. Ni-NTA purification and subsequent SDS-PAGE analysis produced a clear band for OocP but none for OocQ (Figure S1). This shows that OocP can indeed be purified without a His-tag. Ni-NTA purification after co-expression yielded two distinct bands corresponding to OocP and OocQ, indicating binding of OocQ, which is pulled down along with OocP during affinity-based purification (Figure 1D). We suspected that the interaction between OocP and OocQ could be an interesting handle to introduce hybrid PKS parts to the oocydin *trans*-AT PKS. Due to the unique nature of OocQ, the docking between OocP and OocQ likely leads to selective formation of an OocPQ protein complex. Through this interaction, downstream hybrid PKS module parts could be tethered to the native upstream domains (Figure 1E).

OocR as a possible engineering scaffold

In a first attempt to test our planned approach, we partially reconstituted the deleted PKS region in *S. plymuthica* 4Rx13Δ*oocQR* (Figure 1F). To do so, we introduced *oocQ* and its downstream PKS modules 11 and 12 on a pBAD expression plasmid (pBAD-*oocQR*12) into *S. plymuthica* 4Rx13Δ*oocQR* to get *S. plymuthica* 4Rx13Δ*oocQR* pBAD-*oocQR*12, abbreviated as OocQR12 (see Supplementary Information for detailed information on strain and plasmid properties). This partial rescue mutant should reintroduce the chlorination and elongate the intermediates according to PKS modules 11 and 12 in the wild-type assembly line (an overview of the expected compounds and the corresponding retention times is provided in Figure S17 and Table S8, respectively). As negative controls, *S. plymuthica* 4Rx13 with empty pBAD and *S. plymuthica* 4Rx13Δ*oocQR* with empty pBAD were analyzed (Supplementary Information). As expected, no ions corresponding to chlorinated or elongated intermediates were detected in the negative controls. While traces of a released chlorinated compound (**1**) were found in OocQR12, only trace amounts of putative elongated intermediates were found in UHPLC-HRMS analysis of extracts of OocQR12 (Figures 2A-B and S18). Since we have shown previously that intermediates are released throughout the assembly line in *S. plymuthica* 4Rx13 strains that lack the terminal condensation domain *oocS*$_C$,[19] we tried to restore the processivity naturally present in a PKS by reintroducing the offloading domain. To facilitate release of the final intermediate, we cloned native *oocS*-C downstream of *oocR*-KS12 onto pBAD-*oocQR*12 (Figure 1F). Next, *S. plymuthica* 4Rx13Δ*oocQR* was complemented with the resulting plasmid pBAD-*oocQR*12-*oocS*-C to obtain OocQR12-C (Figures 1F and 2A). Alternatively, we added the releasing C domain from the basiliskamide *trans*-AT PKS in OocQR12-BasC (Figure 1F). In the latter, no released product was observed (Figures 2A-B and S18), while UHPLC-HRMS analysis of the extracted expression culture of the OocQR12-C strain provided ions corresponding to the halogenated and elongated intermediates of modules 11 and 12, which were expected based on the biosynthetic model (Figures 2B-C, and S17). In addition to released chlorinated intermediate observed as **1**, ions corresponding to **1** with an additional C_2H_2O were observed as **2lin**. This modification is in line with an extension by the introduced module 12, suggesting proper transfer of the intermediate from the first part of the oocydin assembly line to the introduced PKS parts. Further, a congener lacking H_2O relative to **2lin** was detected, possibly indicating cyclization of **2lin** to **2cyc** by the introduced native condensation domain (Figure 2A). The observation

of these masses indicated successful interaction of OocP with the reintroduced PKS proteins OocQ and the starting domains of OocR and subsequent halogenation. With the offloading of the chlorinated, truncated intermediates and the functional interactions between OocP, OocQ and OocR, OocR provides a promising scaffold to unravel design rules for engineering of *trans*-AT PKS assembly lines.
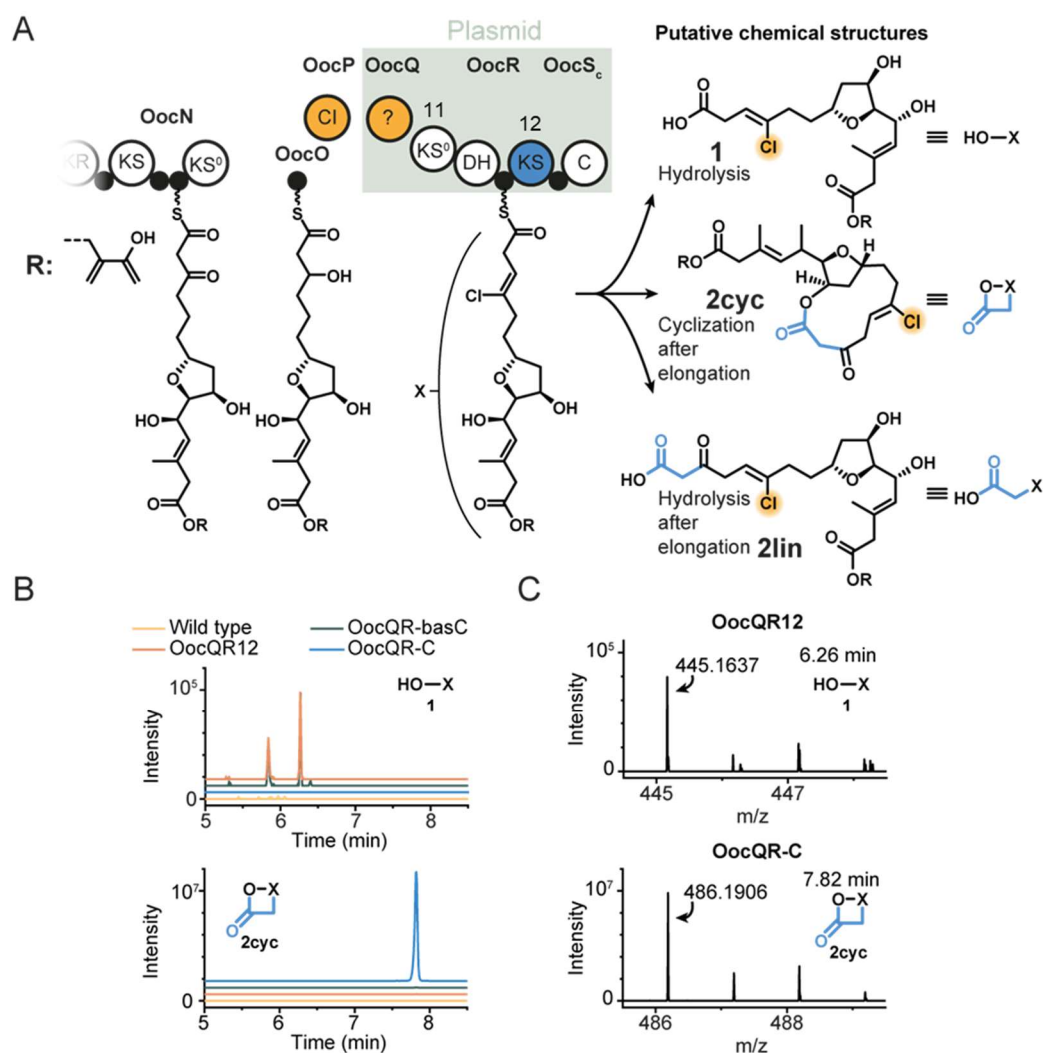
Figure 2. Complementation of *S. plymuthica* 4Rx13 Δ*oocQR* with a plasmid harboring the genes encoding OocQ, the first domains KS⁰-DH-ACP-KS of OocR and the releasing ACP-C domains from OocS yields extended and halogenated oocydin congeners after release by *oocS*-C. A) Setup for the complementation. In a functional hybrid, the blue portion on the hypothetical structures would be introduced by the KS colored in blue. Numbers above KSs indicate their position in the PKS. For easier representation, abbreviations for the structures **1**, **2cyc**, and **2lin** are defined. B) Extracted ion chromatograms for **1** and **2cyc** for the various *S. plymuthica* 4Rx13Δ*oocQR* mutants. C) Isotope patterns of masses attributed to **1** (calculated for $[M+NH_4]^+$: 445.1624, observed 445.1633 and 445.1637) and **2cyc** (calculated for $[M+NH_4]^+$: 486.1889, observed: 486.1906). The data are obtained from the pBAD-*oocQR*12 and pBAD-*oocQR*12-*oocS*-C strains, respectively.

Statistical analysis of residue co-evolution identifies the LPTYPFx5W fusion site

For the construction of hybrid *trans*-AT PKS, we next planned to graft foreign modules downstream of *oocR*-KS12 on plasmid pBAD-*oocQR*12-C. By introducing foreign modules downstream of the second KS domain of OocR, the grafted modules are introduced into a more canonical *trans*-AT PKS region, namely a KS–KR interface (Figure 1E). Using this rationale, we aimed to engineer hybrid oocydin-based *trans*-AT PKS assembly lines by complementing *S. plymuthica* 4Rx13 Δ*oocQR* with plasmids containing the genes for various non-native domains fused between *oocQR*12 and *oocS*-C (Figure 1E).

To guide our experimental recombineering experiments, we first wished to obtain better insights into suitable fusion sites that might yield functional hybrid assembly lines. We have previously reported that natural *trans*-AT PKSs tend to form hybrids by recombining in gene regions immediately downstream of encoded KS domains. This correlates with the general phylogenetic pattern of KSs from *trans*-AT systems, which co-evolve with the modifying domains of the upstream module.[15c, 16] Such hybrids would preserve the match between a substrate-selective KS and the modification installed upstream into the polyketide intermediate. Thus, a fusion site downstream of the KS that maintains the connection between modifying domains and the acceptor KS might be important for preserving the functionality of the assembly line.[15c] Adopting such an evolution-based rationale for PKS engineering, we next aligned protein sequences of KSs and flanking domains to pinpoint post-KS regions of natural recombination more precisely. At their C-terminus downstream of the active-site HGTGT and NIGH motifs (Figure S4), KSs harbor several stretches of highly conserved residues, including a NAHVILEE and a LPTYPFx5W motif. The latter is located at the end of a C-terminal region, termed flanking subdomain, that is thought to play an important role in protein-protein interactions of *trans*-AT PKSs.[20] Based on protein similarities, our alignments using natural hybrid and non-hybrid PKSs suggested that recombinations occurred in the NAHVILEE to LPTYPFx5W region. However, we were unable to further narrow down this region, since even architecturally closely related PKSs exhibited only moderate sequence identities.

More subtle patterns of co-evolution can be uncovered by deep sampling of domain sequences. Such sampling can indicate positions in the sequence that form boundaries of functional units. Statistical coupling analysis (SCA) is well-suited to uncover global patterns of co-evolving positions in multi-domain protein and nucleic acid sequences (Figure 3A).[21] The highly combinatorial nature of *trans*-AT PKS BGCs makes them excellent systems to analyze with SCA. To identify candidate fusion sites, we first constructed and filtered a multiple sequence alignment (MSA) of KS–adapter–KR tridomain (adapter referring to the flanking subdomain) sequences of known *trans*-AT PKS BGCs (see Supplementary Information for full details). Using the filtered and weighted MSA, SCA calculates a covariance matrix that captures the co-evolution of pairs of amino acids at all positions in the MSA (Figure 3B) and identifies groups of co-evolving amino acid positions in the alignment. Through independent component analysis of the covariance matrix, we subsequently clustered correlated groups of strongly covarying positions to obtain co-evolving sectors within the multidomain sequence (Figure 3B and Supplementary Information). As such, the sectors obtained through SCA might identify evolutionary boundaries of the co-evolution of *trans*-AT PKS domains.
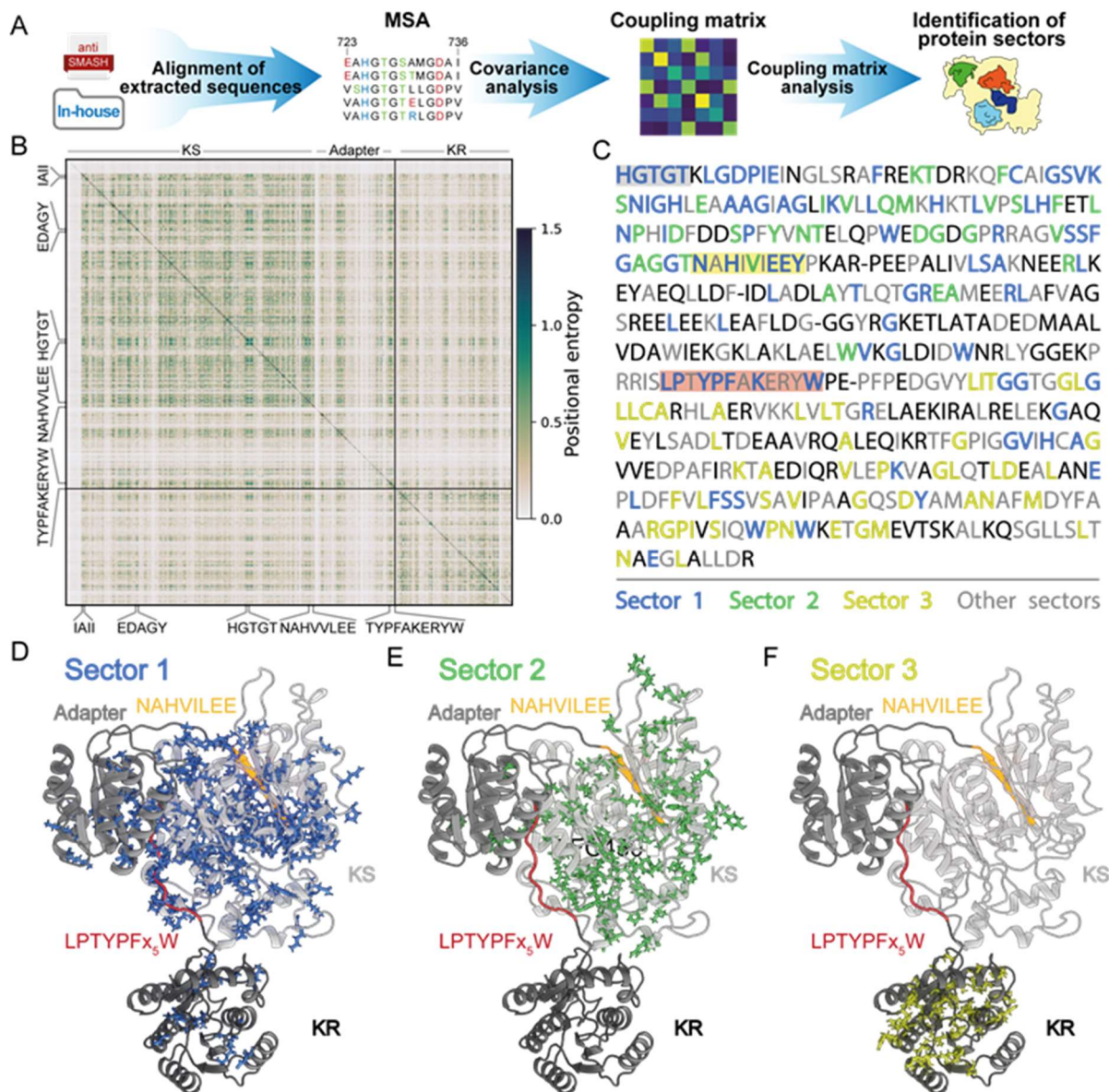
Figure 3. Statistical coupling analysis identifies the LPTYPFx₅W motif as a likely fusion point for hybrid assembly lines. A) Visualization of the statistical coupling analysis workflow on *trans*-AT PKS clusters from the antiSMASH database and our in-house database. The sequences were extracted, aligned and then fed into statistical coupling analysis, which identified co-evolving residues in the proteins as sectors. B) Coupling matrix obtained from the covariance analysis of the filtered MUSCLE MSA of the KS-adapter domain-KR tridomain. The axes show amino acid positions and several amino acid motifs are indicated. C) Color-coded consensus sequence for the C-terminal part of the KS-adapter domain-KR tridomain. Each color represents a sector that corresponds to a subset of co-evolving residues. Residues in sector 1, 2 and 3 are indicated in blue, green and yellow, respectively. Residues in other, minor sectors are indicated in gray and residues that are in no sectors are black. D-F) 3D visualizations of sector 1 (D), sector 2 (E) and sector 3 (F) on a model of the OocR KS-adapter domain-KR tridomain, generated with RoseTTAFold.[22] The KS, flanking subdomain and KR are indicated in light, normal and dark gray, respectively. The yellow and red domains in the ribbon representation indicate the NAHVILEE and LPTYPFx₅W motif, respectively.

The residues in the various sectors obtained from SCA occur in distinctly different parts of the multidomain sequences (Figure 3C). This suggests that the residues in these sectors might be involved in specific functions or interactions of the protein that might be sensitive to disruption. To visualize the distribution of the sectors on the proteins, we computed a structural model of the KS-adapter-KR tridomain using RoseTTAFold[22] and mapped the sectors. Sector 1 (blue in Figure 3C and D) contains highly conserved residues throughout the tridomain such as the IAII motif at the start, the EDAGY motif typical for *trans*-AT PKSs and the conserved HGTGT motif of the KS active site.[23] The presence of these motifs indicates that the positions in sector 1 play a role in general stability and functionality of the enzyme. The residues in sectors 2 and 3 do not include highly conserved residues. Instead, sector 2 (green in Figure 3) is localized on relatively weakly conserved positions in the KS and flanking subdomain and sector 3 (yellow in Figure 3) is localized exclusively on the KR (Figure 3C, E, and F). This clear separation indicates that residues found in sectors 2 and 3 are important for the specific function of the respective enzyme domains. The presence of several positions downstream of the NAHVILEE motif in sector 2 further suggests that the flanking subdomain is intrinsically coupled to and co-evolving with the KS. In hybrid assembly lines that are fused directly downstream of the NAHVILEE motif, interactions between strongly co-evolving amino acid interactions might be disturbed and thereby disrupt functionality. Moreover, inspection of the co-variance matrix (Figure 3B) clearly indicates that the LPTYPFx$_5$W motif at the end of the flanking subdomain functions as a sharp boundary between the two sectors: residues within the KS-AT adapter domains upstream and residues within the KR domain downstream of this motif show strong internal co-evolution, but only marginal co-evolution between these two regions is observed.

A number of control calculations show that the identification of the LPTYPFx$_5$W motif as boundary of the KS domain is not an artifact of the alignment quality, positional conservation or position in the alignment (Figures S6-13). SCA results of MSAs constructed with either Clustal-Omega[24] or MAFFT[25], as well as SCA results of MSAs that were additionally filtered for highly conserved (>80%) positions showed comparable results and also identify the LPTYPFx$_5$W motif as a boundary between independently co-evolving domains (Figure S14). Furthermore, SCA results of MSAs of several other domain motif sequences consistently show strong covariance between residues up- and downstream of the NAHVILEE motif whereas the residues up- and downstream of the LPTYPFx$_5$W motif only show slight covariance (Figure S14). Further, modelling of the two motifs (Figures 3 D-F) suggests that the LPTYPFx$_5$W motif is accessible compared to the NAHVILEE motif which is predicted as a β-sheet buried within the KS and therefore maintains a large number of contacts to neighboring residues. Interestingly, an earlier model with the Phyre2 Protein Fold Recognition Server[26] provided less clear structural predictions with both motifs buried within a protein (Figure S15), which reinforces the necessity of good structural insights for engineering attempts. Our SCA analysis indicates that the LPTYPFx$_5$W motif marks a co-evolutionary boundary between the KS–adapter didomain and the downstream domains and might serve as a promising fusion point in the engineering of *trans*-AT PKS assembly lines.

Experimental verification of fusion sites

To test the computationally predicted LPTYPFx$_5$W as well as the NAHVILEE recombination sites, we constructed hybrid assembly lines for experimental validation using our plasmid-based setup. To minimize the number of interactions between non-native proteins, we started with mutants wherein we added the terminal region from another *trans*-AT PKS after *oocR*-KS12. This way, the hybrid strain only contains a single site of recombination and non-native PKS interface. We selected the terminal modules of the psymberin *trans*-AT PKS from an uncultivated bacterium[23], starting at the ACP upstream of PsyD-KS11. In psymberin biosynthesis, these modules are thought to be responsible for release and aromatic cyclization of the compound.[23] The region encoding these proteins seemed a promising candidate to be introduced downstream of *oocQR*12 because PsyD-KS10 in the module upstream of PsyD-KS11 belongs to a β-keto thioester accepting KS clade. Similarly, OocR-KS12 belongs to a β-keto thioester accepting KS clade. This means that in the hybrid strain, PsyD-KS11 follows a phylogenetically "cognate" KS. The fact that both KSs originate from a similar module environment might facilitate non-native interactions in hybrid *trans*-AT PKSs. The introduced terminal psymberin modules *psyD*-ACP-KS11-?-DH-ACP-TE (? refers to a domain of unknown function) are thought to perform a polyketide elongating resulting in a β-keto group, yielding proposed products **3lin** and **3cyc** (Figure 4A).[23] To test the relevance of the recombination site proposed by SCA, the terminal modules of the psymberin *trans*-AT PKS were cloned downstream of *oocQR*-KS12 using either the NAHVILEE or the LPTYPFx$_5$W motif for artificial fusion (Figure 4B). The resulting plasmids pBAD-*oocQR*12^L-*psyD*$_{term}$ and pBAD-*oocQR*12^N-*psyD*$_{term}$ (superscript "L" and "N" referring to the respective motifs) were introduced into *S. plymuthica* 4Rx13Δ*oocQR* to produce strains OocR12^L-PsyD$_{term}$ and OocR12^N-PsyD$_{term}$. The strains were grown under expression conditions, extracted with ethyl acetate, and analyzed by UHPLC-HRMS. In both strains, traces of **2** corresponding to the released intermediate after chlorination and elongation by the OocQR PKS parts were observed (Figures S19 and 27-28). Gratifyingly, we observed in addition at high intensity a new ion **3cyc** (calculated for [C$_{25}$H$_{31}$ClO$_9$+H]$^+$ 511.1730) corresponding to another KS elongation (+C$_2$H$_2$O) compared to **2cyc**. This compound was only detected in extracts of the strain with the fusion at the LPTYPFx$_5$W motif (Figure 4C), but not in OocR12^N-PsyD$_{term}$. As the terminal module in psymberin naturally leads to aromatic cyclization, we measured the UV spectrum of extracts from OocR12L-PsyD$_{term}$. In the UV spectrum, no peak at wavelengths of aromatic cycles were observed. This suggests that aromatic cyclization of the triketone does not take place (Figure S22). For neither OocR12^N-PsyD$_{term}$ nor OocQR12-C, extracts showed an ion corresponding to **3cyc**, suggesting that an LPTYPFx$_5$W-fused KS domain is required to produce this compound.

Encouraged by the possible products observed in the engineered *trans*-AT PKS assembly line, we studied the influence of the fusion site on the processivity of the assembly line also in an internal hybrid. Compared to the terminal hybrid, an internal hybrid contains an additional fusion site where the non-native PKS part is re-connected to the native releasing domains, in our case the terminal ACP-C domains from OocS. Using all combinations of the NAHVILEE and LPTYPFx$_5$W motif at the fusion points upstream and downstream of the foreign part, we inserted the minimal *lbmD*-ACP-KS12 module from the lobatamide *trans*-AT PKS between *oocR*-KS12 and *oocS*-C (abbreviated as C) on pBAD-*oocQR*12-C (Figure 4D). Transformation of *S. plymuthica* 4Rx13Δ*oocQR* with the created plasmids yielded mutants OocR12^L-LbmD12^L-C, OocR12^L-LbmD12^N-C, OocR12^N-LbmD12^L-C and OocR12^N-LbmD12^N-C. After growth under expression conditions, the cultures were extracted with ethyl acetate and subjected to UHPLC-HRMS analysis (Figures S19 and S29-32). In case of successful elongation over two fusion sites and subsequent cyclic release by *oocS*-C, we expect the ion **3cyc** and for a single elongation, the ion **2cyc**. Out of the four hybrid variants, **3cyc** is observed only in extracts of the OocR12^L-LbmD12^L-C. The ion corresponding to the singly elongated **2cyc** is observed most strongly in the OocR12^L-LbmD12^N-C extract and with less intensity in the OocR12^L-LbmD12^L-C extract (Figure

4E). The retention time of each ion is equal to the retention time observed for the same ion in the OocR12^L-PsyD$_{term}$ extract. This strongly suggests that the molecular structures of these compounds are the same. For the OocR12^N-LbmD12^N-C and OocR12^N-LbmD12^L-C extracts, only the ion corresponding to the hydrolyzed intermediate **1** were observed (Figure S19). No masses corresponding to compounds **2** or **3** were observed in these extracts, suggesting that fusion after the NAHVILEE motif does not yield functional hybrid assembly lines. The results for the OocR12-LbmD12-C hybrids thereby show that masses corresponding to compounds modified by domains downstream of the fusion site are only observed when the hybrid assembly lines are fused at the LPTYPFx$_5$W motif. To elucidate the structure and test whether the compound observed at 6.2 minutes with m/z = 511.1730 is the proposed **3cyc**, we cultured the OocR12^L-LbmD12^L-C on a larger scale and extracted the culture with ethyl acetate. Preparative HPLC separation for NMR analysis is currently under way.

These results indicate that the native flanking subdomain region is important in maintaining processivity of the assembly line. As such, the LPTYPFx$_5$W fusion site appears to yield functional hybrid assembly lines in which PKS parts from different *trans*-AT PKSs operate in concert.
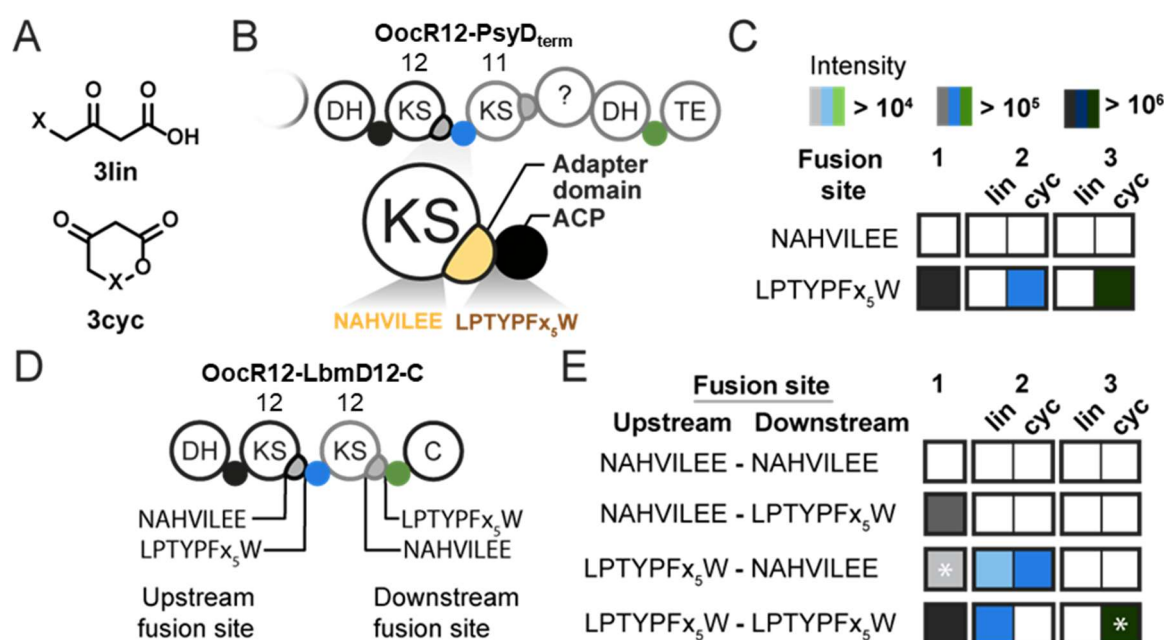
Figure 4. Assembly lines fused after the LPTYPFx$_5$W motif produce hybrid products. A) Putative chemical structures of the products after non-native KS elongation. Hydrolysis of the product from the ACP leads to the linear **3lin**, whereas cyclization by the condensation domain gives the putatively cyclized **3cyc** (loss of one H$_2$O compared to **3lin**). B) Domain architecture of the pBAD-*oocQR*12-*psyD*$_{term}$ hybrids, which are fused after either the NAHVILEE or LPTYPFx$_5$W motif. Numbers above KSs indicate their position in the PKS. C) Heat maps indicating the intensity of the observed peaks in the UHPLC-HRMS analysis of chlorinated masses corresponding to the ammonium adducts of various congeners. The intensity color code is used for all heatmaps. D) Domain architecture of the OocR12-LbmD12-C hybrids, with the alternative upstream and downstream fusion points indicated. E) Heat maps indicating the intensity of the observed peaks in the UHPLC-HRMS analysis of chlorinated masses corresponding to the ammonium adducts of various congeners. Asterisk indicates that the proton adduct was observed.

Grafting of non-cognate KSs impairs the production of extended oocydin hybrids

With the LPTYPFx$_5$W motif identified as fusion site yielding functional terminal and internal hybrids, we investigated the scope and limitations of adding further module types into the oocydin assembly line. In the first set, we use modules truncated to the minimal ACP-KS domains. The KS in such a truncated module is non-cognate in the sense that it must accept an incoming intermediate that has only undergone partial β-keto processing compared to the native intermediate. We include these non-cognate KSs to test, whether the functionality is preserved if a KS is confronted with a non-native intermediate after domains are removed and thus contact points with the KS altered.

To interrogate whether the integrity of foreign modules need to be preserved, we first designed two mutants in which a minimal module, ACP-KS, stemming from larger upstream modules, is inserted between *OocR*-KS12 and *OocS*-C. For this, we selected ACP-KS11 from the lobatamide *trans*-AT PKS[27] and ACP-KS5 from the bacillaene PKS[28]. In their native PKSs, these didomains are part of modules with upstream DH and KR domains (Figure 5B), i.e., the downstream KSs accept α,β-unsaturated instead of β-keto intermediates. The ACP-KS didomains were cloned onto pBAD-*oocQR12*-*oocS*-C by using the LPTYPFx$_5$W motifs as fusion sites and *S. plymuthica* Δ*oocQR* was transformed with the resulting plasmids to yield strains OocR12-LbmD11-C and OocR12-PksL5-C. After cultivation under expression conditions, UHPLC-HRMS analysis of the extracts showed the presence of **3cyc** (Figures 5C, S20 and S33-34). Although the intensity of the peaks observed for **3cyc** was 10-100 fold lower than in extracts of OocR12^L-LbmD12^L-C strains, this indicates that the truncated, minimal module is able to extend the polyketide intermediate despite the lack of the native downstream modifying enzymes.

In a third hybrid assembly line, the section of the gene *pelD* encoding KS9 of the peloruside BGC[29] was inserted between *oocR*12 and *oocS*-C on pBAD-*oocQR*12-C to yield the strain OocR12-PelD9-C. Even though PelD-KS9 is part of a minimal ACP-KS module, it is part of a phylogenetic clade of KSs that accept a range of reduced substrates instead of the β-keto thioester expected based on the biosynthetic model.[29] UHPLC-HRMS analysis of OocR12-PelD9-C extracts only showed trace amounts of ions that may tentatively be assigned to **3** (Figure S36). In a second hybrid with an incomplete native module, we targeted module 13 of the *lcn trans*-AT PKS. This module is as such minimal consisting of in-line ACP and KS, but according to the biosynthetic model, a KR acts on the module *in trans*. Congruously, LcnB-KS13 clades with KSs accepting β-hydroxy-containing intermediates and not β-keto as would be expected based on the module architecture. As for the previous mutants, we inserted the respective ACP-KS didomain onto pBAD-*oocQR*12-*oocS*-C with the LPTYPFx$_5$W fusion sites and introduced the plasmid into *S. plymuthica* Δ*oocQR* to obtain the strain OocR12-LcnB13-C. UHPLC-HRMS analysis of culture extracts revealed two products, **2lin** and **2cyc** (Figure 5). However, **3lin** and **3cyc** expected as products after elongation by LcnB13 were absent, indicating impaired KS functionality without the *trans*-acting KR. Thus, compared to hybrids containing minimal modules in the native PKS, the processivity of those carrying truncated minimal modules is impaired. This observation may be explained by the disruption of protein-protein interactions around the KS. Based on the assumption, that KSs from *trans*-AT PKSs co-evolve with their upstream domains, such truncated modules lead to non-functionality as the KS is confronted with a non-native set of domains and, consequently, with a non-native intermediate.
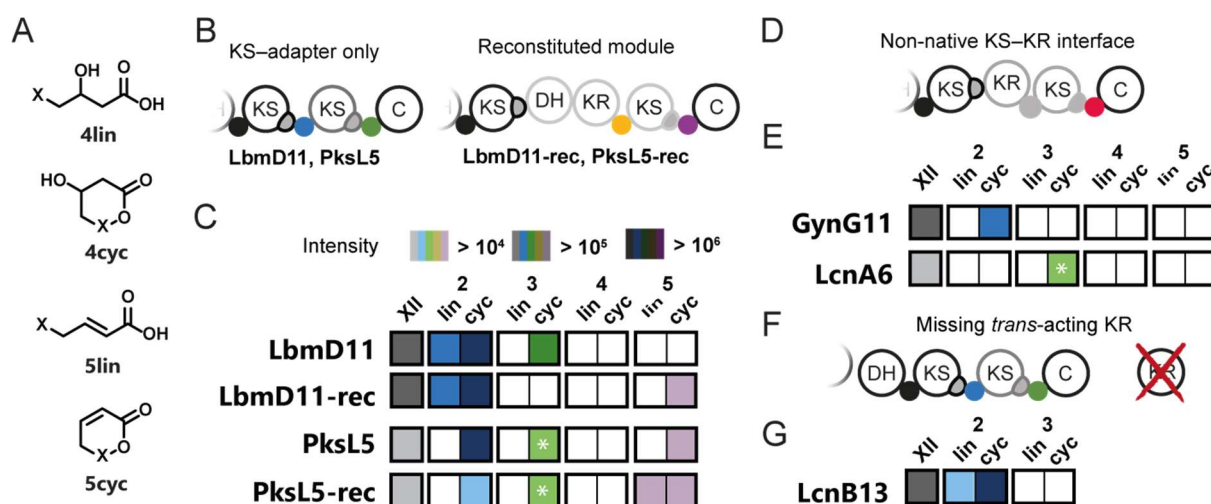
Figure 5. A range of domains can be inserted into hybrid assembly lines. A) Putative chemical structures of the linear and cyclized products **4lin** and **4cyc**, and **5lin** and **5cyc**, which might be produced in hybrid assembly lines containing inserted KR-ACP-KS and DH-KR-ACP-KS domains, respectively. B) Architectures of the OocR12-Lbm11-C and OocR12-PksL5-C truncated module and reconstituted module hybrids, respectively. The introduced PKS parts are marked in grey. C) Heatmaps showing the intensity of chlorinated ions corresponding to NH_4^+-adducts of the various congeners of **1**, **2**, **3**, **4**, and **5** in extracts of OocR12-LbmD11-C, OocR12-LbmD11$_{rec}$-C, OocR12-PksL5-C, and OocR12-PksL5$_{rec}$-C. Fields with an asterisk show the intensity of the proton adduct. D) Architecture of the OocR12-GynG11-C and OocR12-LcnA6-C hybrids. The introduced PKS parts are marked in grey. E) Heatmaps showing the intensity of chlorinated ions corresponding to NH_4^+-adducts of the various congeners of **1**, **2**, **3**, **4**, and **5** in extracts of OocR12-GynG11-C and OocR12-LcnA6-C. Fields with an asterisk show the intensity of the proton adduct. F) Architecture of the OocR12-LcnB13-C hybrid. The introduced PKS part is marked in grey. The *trans*-acting KR from the *lcn* biosynthetic pathway (LcnG) was not included and is crossed out. G) Heatmaps showing the intensity of chlorinated ions corresponding to NH_4^+-adducts of the various congeners of **1**, **2**, and **3** in extracts of OocR12-LcnB13-C.

Hybrids containing larger non-native modules are partially functional

Encouraged by the successful elongation of the polyketide intermediates by the truncated minimal modules, we also genetically inserted the complete LbmD11 and PksL5 modules between *oocR*-KS12 and *oocS*-C (Figure 5B, full modules are termed LbmD11$_{rec}$ and PksL5$_{rec}$). Both modules naturally consist of a DH-KR-ACP-KS domain series, which would lead to compound **5cyc** if fully functional (Figure 5A). A partially modified released intermediate is expected to be present as **4cyc**. In case of a non-cyclizing release, **4lin** or **5lin** would be expected as provided in Figure 5A. As for the minimal module mutants, we exclusively used the LPTYPFx$_5$W motif in this set of extended module hybrids. After cloning the modules into pBAD-*oocQR*12-*oocS*-C and transformation into *S. plymuthica* Δ*oocQR* to yield strains OocR12-LbmD11$_{rec}$-C and OocR12-PksL5$_{rec}$-C, strain cultures were extracted. UHPLC-HRMS analysis (Figures S20 and 36-37) indicated that both strains produced new compounds with masses corresponding to dehydrated **5** (Figure 5C). Although both extracts still showed the presence of intermediates **1**, **2lin** and **2cyc**, which are not modified by the domains downstream of the fusion site, and of the KS-modified **3cyc**, the signals of these compounds are slightly less intense than observed in the truncated hybrids. Furthermore, ions corresponding to **4** are only present at the same retention time as the ions observed as **5**. No additional intermediates that may be attributed to keto-hydroxy extended products **4** were found with certainty. This suggests that the DH domain retains full functionality despite its non-native KS-DH interface and very efficiently dehydrates the incoming hydroxylated intermediate. This efficient processing indicates that *trans*-AT PKS hybrid assembly lines fused at the LPTYPFx$_5$W

motif can accommodate both truncated and extended modules that incorporate additional modifying domains besides the KS, such DH and KR domains.

Having shown that our engineering strategy can produce functional assembly lines with non-native KS-ACP or KS-DH boundaries, we constructed a next set of hybrids in which a KR domain is fused directly downstream of the KS, similar to the module architecture in wild-type *S. plymuthica* 4Rx13. For this, we inserted the KR-ACP-KS domains from the gynuellalide (module GynG11) and lacunalide (LcnA6) cluster, respectively, between OocR12 and OocS-C. After construction of the plasmids, transformation of *S. plymuthica* Δ*oocQR* yielded strains OocR12-GynG11-C and OocR12-LcnA6-C. The OocR12-GynG11-C extract obtained after strain cultivation showed the expected hydroxylated product, **4cyc** at low abundance (Figure S21). In the extract of OocR12-LcnA6-C, no ions corresponding to hydroxylated products were detected (Figure S21). We hypothesized that the near absence of processivity in these hybrid assembly lines might be due to the α-substitution of the native substrates of GynG-KS11 and LcnA-KS6. This specificity of the GynG11 and LcnA6 KSs to process α-substituted substrates contrasts with the processivity of the other hybrid assembly lines, in which the LbmD11 and PksL5 KSs were also confronted with polyketide intermediates different from their native substrates. In these latter cases, however, both native intermediates are unbranched, α,β-unsaturated ketones. This discrepancy may indicate that the nature of the native KS intermediate determines to which extent non-native substrates are elongated by the hybrid assembly lines. Despite the absence of hybrid products in the OocR12-LcnA6-C and the OocR12-GynG11-C hybrid assembly lines, our results clearly show that LPTYPFx$_5$W fusion is in many cases a successful strategy to generate functional *trans*-AT PKS hybrids.

Construction of terminal hybrids in bacillaene

To test the general applicability of *trans*-AT PKS recombineering, we attempted to apply our strategy to another model system. We chose the bacillaene (*pksX* or *bae*) PKS from the genetically accessible *Bacillus subtilis* (*B. subtilis*), one of the best-characterized *trans*-AT systems.[28, 30] As *B. subtilis* can be modified by natural transformation, we generated hybrid PKSs directly on the *B. subtilis* genome by double cross-over. We fused different lengths of the terminal PKS region from the psymberin PKS (*psyD*-ACP-KS11-?-DH-ACP-TE as in OocR-PsyD$_{term}$) at either the LPTYPFx$_5$W or NAHVILEE motif (Figure S44). Integration into the target regions of the bacillaene biosynthetic gene cluster in *B. subtilis* DK1042 yielded *B. subtilis* PksM8^N-PsyD10, PksM8^L-PsyD10, PksL4^N-PsyD10, PksL4^L-PsyD10, and PksL4^L-PsyD11 (see the Supplementary Information for a full overview of the mutants). For analysis, the strains were grown at bacillaene producing conditions and culture extracts were analysed by UHPLC-HRMS. For none of the hybrids, however, masses corresponding to the expected modified polyketide fragments were observed. Instead, intermediates up to the PKS fusion point accumulated, indicating stalling of the assembly line due to impaired function of the downstream foreign module.

Conclusion and Outlook

Recombination of PKSs to obtain polyketides with new chemical scaffold has been a challenge. Engineering efforts in *cis*-AT PKSs have produced hybrid assembly lines for minimal triketides or compounds differently reduced compared to the native polyketide.[11] But so far, larger module exchanges to produce new polyketide scaffolds have remained rare.[31] For the biochemically more complex *trans*-AT PKSs, no successful modular engineering attempts have been reported. In this work, we exploited the natural propensity of *trans*-AT PKSs to form hybrids between different assembly lines as a basis for engineering.[15b] This natural mosaic-like evolution and the propensity to integrate numerous biochemically distinct functions[7, 19] hints at an increased tolerance of this enzyme family towards foreign modules. To utilize this property, general guidelines have to be investigated first. Using the unusual oocydin *trans*-AT PKS in *S. plymuthica* as a model system, we computationally and experimentally evaluated fusion sites to engineer first sets of functional chimeric *trans*-AT PKSs.

For a flexible experimental setup, we relied on an unusual protein-protein interaction in the biochemically versatile *trans*-AT PKS for oocydin. We created a *S. plymuthica* 4Rx13 Δ*oocQR* strain, wherein the *ooc trans*-AT PKS was truncated within the halogenating enzyme pair and successfully complemented the strain with the first two downstream modules. This indicated that the affinities between the halogenating enzyme pair OocPQ could be used as a handle to introduce new modules to the native host in a plasmid-based way to generate products that can be readily detected based on their chlorination. Co-evolution analysis predicted a candidate fusion site in the KS flanking subdomain for which disruption of important domain interactions would be minimized. The subsequent experimental confirmation of this site downstream of KSs, rather than ACPs as the traditionally defined module boundary, is in line with phylogenetic analyses showing that the phylogeny of KSs from *trans*-AT systems tightly correlates with the modifications installed by the upstream module.[15b, 15c, 16] It also matches the observation that many natural hybrids of *trans*-AT PKSs recombine at gene positions downstream of encoded KSs.[15b, 18, 23] Evolutionary adjustments toward a specific intermediate might be a consequence of important domain contacts in this PKS region and/or of KSs being substrate-selective and acting as gatekeepers to ensure that chain extension is only performed when upstream modifications have been completed.[15c] In our studies, the engineered PKSs were only active if the LPTYPFx$_5$W motif was used, independent of whether the downstream module was extending or releasing.

This was in clear contrast to the inactive NAHVILEE fusions at the KS-flanking subdomain boundary, but matches the predictions derived from computational coevolution analyses. In a structural model, the flanking subdomain closely interacts with the KS and the LPTYPFx$_5$W motif is found as an exposed linker region on the periphery of the KS. Interestingly, the LPTYPFx$_5$W motif and the flanking subdomain is also conserved in *cis*-AT PKS, where it is found downstream of the AT in each module. It is worth investigating whether this subdomain may serve as a more general strategic site in PKS engineering.

Our results further corroborate the *trans*-AT PKS correlation rule. In grafting experiments with minimal modules, an introduced KS that naturally recognizes intermediates with β-keto functional groups yielded better results than non-matching KSs from truncated modules. This observation underlines the importance of matching KSs, of which at least some have been shown to exhibit high substrate selectivity and might therefore serve as gatekeepers.[32] In an attempt to recover activity of the non-matching minimal modules, their missing module parts were introduced as well to create larger module hybrids. For these mutants, only traces of further modified products were observed. Apart from the possible biochemical reasons already mentioned, the increased size of the plasmid may pose problems. Broad host-range plasmids designed for the expression of larger-sized constructs as well as cloning

strategies involving the direct integration of the foreign PKS parts into the host genome could help overcome this bottleneck.

The suitability of the engineering motif for general *trans*-AT PKS engineering was probed with the bacillaene PKS as a second model system. In contrast to plasmid-based engineering applied to oocydin, the non-native PKS parts were directly inserted into the genome of *B. subtilis*. Unfortunately, for none of the hybrid constructs, transfer of intermediates from the native to the non-native PKS parts was observed. In addition to evolutionary and statistical insights, structural biology is another key player in deciphering polyketide biosynthesis. Since maintaining protein-protein interactions in engineered assembly lines is important, it is conceivable that the introduction of multiple domains into a non-native context disrupts proper protein-protein interactions. Prolonged growth in a directed evolutionary setting in which polyketide production is enforced may help overcome this problem. Random mutations along with selective pressure could improve protein-protein interactions in these engineered systems. Still, PKSs are highly complex involving a large number of protein interfaces. In PKSs, the substrate is channeled through a series of modules while connected to ACPs. Throughout this journey, the correct orientation of domains within the assembly line is necessary to allow for interactions between the ACP and the enzymatic domains as well as for shuttling of the ACP-bound substrate to the downstream module.[17b] Recent structural studies have indicated that the KSs in *cis*-AT PKSs act in a turnstile mechanism that ensures that intermediates on ACPs are only transferred to downstream modules and not backwards.[33] The accessibility of the KS active site is determined by the positioning of the AT.[17b] *Trans*-AT PKSs likely rely on a similar mechanism. As they lack module-encoded ATs, the remnant flanking subdomain region may take over the function of attaining the closed turnstile conformation. In bacillaene, it is assumed, that the flanking subdomain further helps align PKSs in parallel to form a gigantic raft.[20] While it is not known whether all *trans*-AT PKSs form such superstructures, non-native proteins may have different symmetries and can thus not fit into the preexisting inter-PKS superstructures adopted by the beginning of the PKS. This might have a detrimental effect on PKS engineering in the bacillaene biosynthetic pathway. Another factor to take into account is the connectivity between PKS parts that are not encoded on a single protein. In *cis*-AT PKSs, modules are usually connected via a conserved set of docking domains at the ACP-KS interface if the PKS spans over multiple proteins.[34] In *trans*-AT PKSs, split-extension modules are more commonly observed and so far only a number of specific docking sites have been studied.[32b, 34-35] However, due to the high number of unusual module architectures and domains acting *in trans*,[7] a large part of these docking interactions remain to be studied.

In general, many factors contribute to the success or failure of a hybrid *trans*-AT PKS assembly line. We show that the choice of fusion site is important for proper transfer of intermediates between non-native PKS parts and statistically corroborate that the flanking subdomain region found downstream of most *trans*-AT PKS assembly lines should stay with its cognate KS. Further, KSs should be replanted along with all enzymes acting in its module - including enzymes acting *in trans*, that may be difficult to spot. It is likely that not a single flanking subdomain, but a complex interplay between multiple enzymes decides about successful protein docking. While advances in structural biology will surely help elucidating the complex interplay between components of PKS megaenzymes, our results provide first insights into prerequisites for hybrid *trans*-AT PKS production.

Acknowledgements

References

[1]   D. J. Newman, G. M. Cragg, K. M. Snader, *Nat. Prod. Rep.* **2000**, *17*, 215-234.

[2]   a) R. E. Nelson, K. M. Hatfield, H. Wolford, M. H. Samore, R. D. Scott, II, S. C. Reddy, B. Olubajo, P. Paul, J. A. Jernigan, J. Baggs, *Clinical Infectious Diseases* **2021**, *72*, S17-S26; b) E. F. S. Authority, E. C. f. D. Prevention, Control, *EFSA Journal* **2021**, *19*, e06490.

[3]   J. Clardy, C. Walsh, *Nature* **2004**, *432*, 829-837.

[4]   a) K. J. Weissman, *Nat. Prod. Rep.* **2016**, *33*, 203-230; b) K. J. Weissman, P. F. Leadlay, *Nat. Rev. Microbiol.* **2005**, *3*, 925-936.

[5]   a) G. A. Hudson, D. A. Mitchell, *Curr. Opin. Microbiol.* **2018**, *45*, 61-69; b) M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. Van Heel, J. H. Viel, V. Bandarian, E. Dittmann, O. Genilloud, Y. Goto, *Nat. Prod. Rep.* **2021**, *38*, 130-239; c) A. S. Brown, M. J. Calcott, J. G. Owen, D. F. Ackerley, *Nat. Prod. Rep.* **2018**, *35*, 1210-1228; d) J. F. Barajas, J. M. Blake-Hedges, C. B. Bailey, S. Curran, J. D. Keasling, *Synth. Syst. Biotechnol.* **2017**, *2*, 147-166.

[6]   J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380-416.

[7]   E. J. Helfrich, J. Piel, *Nat. Prod. Rep.* **2016**, *33*, 231-316.

[8]   C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688-4716.

[9]   E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, *Nat Chem Biol* **2019**, *15*, 813-821.

[10]   E. Kalkreuter, J. M. CroweTipton, A. N. Lowell, D. H. Sherman, G. J. Williams, *J. Am. Chem. Soc.* **2019**, *141*, 1961-1969.

[11]   A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, *Chem. Rev.* **2019**, *119*, 12524-12547.

[12]   a) J.-M. Massicard, C. Soligot, K. J. Weissman, C. Jacob, *ChemComm* **2020**, *56*, 12749-12752; b) T. Miyazawa, M. Hirsch, Z. Zhang, A. T. Keatinge-Clay, *Nat. Commun.* **2020**, *11*, 1-7.

[13]   T. Kornfuehrer, A. S. Eustáquio, *MedChemComm* **2019**, *10*, 1256-1272.

[14]   R. McDaniel, C. M. Kao, S. J. Hwang, C. Khosla, *Chem. Biol.* **1997**, *4*, 667-674.

[15]   a) E. J. Helfrich, J. Piel, *Nat Prod Rep* **2016**, *33*, 231-316; b) E. J. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, *Nat. Commun.* **2021**, *12*, 1-14; c) E. J. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, *Nat. Chem. Biol.* **2019**, *15*, 813-821.

[16]   T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225-233.

[17]   a) S. R. Bagde, I. I. Mathews, J. C. Fromme, C.-Y. Kim, *Science* **2021**, *374*, 723-729; b) D. P. Cogan, K. Zhang, X. Li, S. Li, G. D. Pintilie, S.-H. Roh, C. S. Craik, W. Chiu, C. Khosla, *Science* **2021**, *374*, 729-734.

[18]   R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, *Nat. Chem. Biol.* **2015**, *11*, 705-712.

[19]   F. Hemmerling, R. A. Meoded, A. E. Fraley, H. A. Minas, C. L. Dieterich, M. Rust, R. Ueoka, K. Jensen, E. J. Helfrich, C. Bergande, M. Biedermann, N. Magnus, B. Piechulla, J. Piel, *Angew. Chem. Int. Ed.* **2022**.

[20]   D. C. Gay, D. T. Wagner, J. L. Meinke, C. E. Zogzas, G. R. Gay, A. T. Keatinge-Clay, *J. Struct. Biol.* **2016**, *193*, 196-205.

[21]   a) S. W. Lockless, R. Ranganathan, *Science* **1999**, *286*, 295-299; b) O. Rivoire, K. A. Reynolds, R. Ranganathan, *PLoS Comput. Biol.* **2016**, *12*, e1004817; c) A. S. Walker, W. P. Russ, R. Ranganathan, A. Schepartz, *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 19879-19887.

[22]   M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, *Science* **2021**, *373*, 871-876.

[23]   K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews, J. Piel, *Nat. Chem. Biol.* **2009**, *5*, 494-501.

[24]   F. Sievers, D. G. Higgins, in *Multiple Sequence Alignment*, Springer, **2021**, pp. 3-16.

[25]   K. Katoh, K.-i. Kuma, H. Toh, T. Miyata, *Nucleic Acids Res.* **2005**, *33*, 511-518.

[26]   L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**, *10*, 845-858.

[27]   R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem., Int. Ed.* **2020**, *59*, 7761-7765.

[28]   J. Moldenhauer, X. H. Chen, R. Borriss, J. Piel, *Angew. Chem. Int. Ed.* **2007**, *46*, 8195-8197.

[29]   M. Rust, E. J. Helfrich, M. F. Freeman, P. Nanudorn, C. M. Field, C. Rückert, T. Kündig, M. J. Page, V. L. Webb, J. Kalinowski, *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 9508-9518.

[30]   C. Scotti, M. Piatti, A. Cuzzoni, P. Perani, A. Tognoni, G. Grandi, A. Galizzi, A. M. Albertini, *Gene* **1993**, *130*, 65-71.

[31]   a) X. Chen, P. Wei, L. Fan, D. Yang, X. Zhu, W. Shen, Z. Xu, P. Cen, *Appl. Microbiol. Biotechnol.* **2009**, *83*, 507-512; b) L. Su, L. Hotel, C. Paris, A. Brachmann, J. Piel, C. Jacob, B. Aigle, K. Weissman, **2021**.

[32]   a) Darren C. Gay, G. Gay, Abram J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, Neil J. Oldham, J. Piel, Adrian T. Keatinge-Clay, *Structure* **2014**, *22*, 444-451; b) M. Jenner, S. Kosol, D. Griffiths, P. Prasongpholchai, L. Manzi, A. S. Barrow, J. E. Moses, N. J. Oldham, J. R. Lewandowski, G. L. Challis, *Nat. Chem. Biol.* **2018**, *14*, 270-275.

[33]   B. Lowry, X. Li, T. Robbins, D. E. Cane, C. Khosla, *ACS Cent. Sci.* **2016**, *2*, 14-20.

[34]   K. J. Weissman, R. Müller, *ChemBioChem* **2008**, *9*, 826-848.

[35]   J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2016**, *11*, 2466-2474.

Experimental Procedures

General
All restriction endonucleases used were obtained from New England BioLabs (Ipswich, MA, USA). For Gibson assembly, the Gibson assembly kit from New England BioLabs (Ipswich, MA, USA) was used. T4 ligase and buffer for GG cloning were obtained from Promega (Fitchburg, WI, USA). Colony PCR was performed with Phusion polymerase (Thermo Fisher Scientific, Waltham, MA, USA) according to manufacturer's instructions. DNA sequencing was conducted by GATC Biotech (Konstanz, Germany) or Microsynth (Balgach, Switzerland). The pJET1.2/blunt and pCR-Blunt II-TOPO cloning kits were purchased from Thermo Fisher Scientific (Waltham, MA, USA) and the pGEM-T Easy subcloning kit was obtained from Promega (Fitchburg, WI, USA). For sonication, a Sonicator Q700 from QSonica, Newton was used and Ni-NTA agarose for protein purification was purchased from Macherey-Nagel. Gel purification and plasmid extraction kits were purchased from Macherey-Nagel.

Sequence alignments of KS sequences
Amino acid sequences of 1092 KS with and without downstream adapter regions were extracted from an in-house database of 88 annotated *trans*-AT PKS clusters. As an outgroup, KS3 and KS5 from the erythromycin *cis*-AT PKS were used. The sequences were aligned using the MUSCLE algorithm with default settings[1] and a phylogenetic tree was computed with FastTree (version 2.1.10 +SSE3 +OpenMP, 16 threads, default settings)[2]. A sequence logo was created using weblogo.[3]

Strains and Culture conditions
B. subtilis, *E. coli*, and *S. plymuthica* strains used in this study are listed in Table S1. We used *B. subtilis* DK1042 which is a derivative of the *B. subtilis* 3610 wild-type strain with increased competence due to a single point mutation in the *comI* gene negatively regulating natural competence.[4] Strains were grown in LB liquid medium and on LB agar plates at 37 °C. To select plasmids in *E. coli* and *S. plymuthica*, antibiotics were used at final concentrations as follows: Ampicillin 100 μg/mL, chloramphenicol 12.5 μg/mL, gentamycin 20 μg/mL, kanamycin 50 μg/mL, and spectinomycin 50 μg/mL. To select for positive transformants in *B. subtilis*, antibiotics were used at final concentrations of 60 μg/mL spectinomycin and 5 μg/mL chloramphenicol.

Generation of *S. plymuthica* 4Rx13 knockouts
The various *S. plymuthica* 4Rx13 knockout strains were generated as described in Domik, et al. [5], using the primer pairs KO_SOD_c22970_fw and KO_SOD_c22970_rev for *S. plymuthica* 4Rx13Δ*oocR* and KO_SOD_b01030_fwd and KO_SOD_c22970_rev for *S. plymuthica* 4Rx13Δ*oocQR* (Table S2).

Generation of expression plasmids for docking studies of OocP and OocQ
The gene pair *oocP* and *oocQ* was amplified from a *Serratia plymuthica* 4Rx13 colony using the primer pair oocP_NheI_f and oocQ_NotI_r (Table S2). After gel purification, the gene fragment was digested with NheI and NotI and cloned into a pET24-derived vector yielding pET24BH*oocPoocQ*. The plasmid was propagated in *E. coli* DH5α and reisolated prior to introduction into *E. coli* BL21 tuner™ (DE3) competent cells (Novagen). The resulting strain was used for the expression of native OocP and OocQ. For the separate expression of OocP and OocQ, the His-tag was removed by PCR with primers oocP_noHis_F, oocP_noHis_R, and oocQ_noHis_F using primers pCDF-Duet_*oocP* and pCDF-Duet_*oocQ* (Katja Jensen, unpublished), respectively.

Protein expression and purification

TB medium supplemented with appropriate antibiotics was inoculated 1/100 with *E. coli* expression strain overnight cultures. The culture was grown to an OD_{600} of 0.8-1.0 and then cooled on ice for 15 min. Expression was induced by addition of 0.1 mM isopropyl β-ᴅ-1-thiogalactopyranoside (IPTG) and induced cultures were grown for an additional 16-20 h at 16 °C. The cells were harvested by centrifugation and the cell pellets were either frozen for later usage or directly processed. Pellets were resuspended in lysis buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol) and disrupted by sonication (QSonica, Newton). The cell lysate was centrifuged for 20 min at 18,000 × g to remove cell debris and the supernatant was incubated with Ni-NTA agarose (Macherey-Nagel) for 15 min at 4 °C. The resin was transferred to a fretted column, washed once with lysis buffer and finally eluted with elution buffer (50 mM phosphate buffer pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol). Elutions were concentrated with Amicon columns and desalted using PD10 spin columns. Protein expression was assessed by SDS-PAGE.

Cloning of hybrid plasmids

The sequences encoding OocQR$_{KS0-11\_to\_KS12}$ and OocS$_{ACP\_C}$ were amplified from *S. plymuthica* 4Rx13 liquid culture and the sequences encoding grafted PKS fragments were amplified from *Gynuella sunshinyii* YC6258 liquid culture (lacunalide[6], lobatamide[7], gynuellalide[8]), previously isolated metagenomic DNA of *Mycale hentscheli* (peloruside[9]), *Bacillus subtilis* DK1042 liquid culture (bacillaene[10]), *Brevibacillus laterosporus* DSM25 (basiliskamide[11]), or a previously isolated pCC1Fos fosmid (psymberin[12]). An overview of the primers used is provided in Table S3. pBAD/Myc-His was used as backbone and the individual constructs were assembled using extension PCRs and Gibson assembly. Assembled constructs were introduced into *E. coli* DH5α, verified by test digest, colony PCR, or sequencing, and then introduced into *S. plymuthica* 4Rx13 Δ*oocQR*. As controls, empty pBAD/myc-His was introduced into *S. plymuthica* 4Rx13 Δ*oocQR*.

Expression and analysis of *S. plymuthica* 4Rx13 strains

For MS-based analysis of *S. plymuthica* 4Rx13 hybrids, the strains were inoculated 1:100 from overnight cultures and grown in enriched potato dextrose broth (EPB: 24 g/L potato dextrose broth (Difco), 6 g/L bactopeptone, 4 g/L yeast extract, 100 mg/L NaCl) supplemented with appropriate antibiotics for 18-24h at 26 °C and 200 rpm. Expression of the introduced PKS parts was induced by the addition of 0.2% ᴌ-arabinose. The cultures were spun down and the supernatant was extracted with ethyl acetate, dried, resuspended in methanol and subjected to UHPLC-HRMS analysis using a Dionex Ultimate 3000 UPLC system connected to a Thermo QExactive mass spectrometer. A solvent gradient (A = H_2O + 0.1% formic acid and B = acetonitrile + 0.1% formic acid) with B at 5% for 0–2 min, 5–50% for 2–4 min, 50–95% for 4–10 min, and 95% for 10–13 min at a flow rate of 1.0 mL/min) was used on a Phenomenex Kinetex 2.6 µM C18 100A (150 × 4.6 mm) column at 27 °C. The MS was operated in positive ionization mode at a scan range of 150–1500 *m/z*. The spray voltage was set to 3.7 kV and the capillary temperature to 320 °C. Collected data of all MS experiments were analyzed using the Thermo Xcalibur 2.2. software.

Plasmid construction for *Bacillus* constructs

As an acceptor vector, pFusA was used. The plasmid was a gift from Adam Bogdanove & Daniel Voytas (Addgene plasmid #31028). All gene fragments and resistance genes used in Golden Gate (GG) cloning were amplified with primers containing the *Bsa*I restriction sites (Table S4). All PCR fragments were subcloned into pJET1.2/blunt, pCRBlunt II-TOPO or pGEM-T Easy according to manufacturer's instructions. Golden Gate cloning (GG) was performed as published by C. Engler and S. Marillonnet using pFusA as an acceptor vector.[13] See Table S5 for a list of plasmids and information on donor and acceptor vectors for GG assembly. The chloramphenicol (Cm) resistance gene was amplified from pACYC184 (New England BioLabs, Ipswich, MA, USA) and subcloned into pCR-Blunt II-TOPO. The

HyperSpac promoter was amplified from pMF37/pDGICZ and fused to the Cm resistance gene using Gibson assembly. Subcloned Golden Gate pieces and Golden Gate constructs were introduced into *E. coli* DH5α. As a common acceptor plasmid, pFusA with a placeholder kanamycin resistance cassette, the chloramphenicol resistance cassette and a common downstream homology was assembled. For assembly of the hybrid constructs, codon-optimized synthetic DNA was acquired from GenScript (sequences are provided in Table S4). The synthetic DNA was inserted into the common acceptor plasmid pFusA/S/C_KanR by restriction cloning with PmeI, NheI and SpeI. The final assemblies were introduced into *E. coli* DH5α.

Bacillus transformation and screening

The transformation protocol was obtained from http://2012.igem.org. 3 mL of freshly prepared medium A (25% glucose 1 mL, 10x medium A (yeast extract 5g, casamino acids 1g, ddH$_2$O add 450 mL) 4 mL, 10x salt solution (NH$_4$)$_2$SO$_4$ 10g, K$_2$HPO$_4$ 69.8g, KH$_2$PO$_4$ 30g, Na$^+$ citrate 5g, MgSO$_4$·7 H$_2$O 1g, ddH$_2$O add 500 mL) 4.5 mL, H$_2$O ad 45 mL) were inoculated with a single colony from a fresh plate and incubated at 37 °C and 180 rpm overnight. 20 mL of freshly prepared and pre-warmed (37 °C) medium A was inoculated to an OD$_{600}$ of 0.1 with the overnight culture. Cells were grown to an OD$_{600}$ of 1.0 at 37 °C and 180 rpm. 45 mL of prewarmed freshly prepared medium B (25% glucose 1 mL, 10x medium A 4 mL, 10x salt solution 4.5 mL, H$_2$O ad 45 mL, 1 M CaCl$_2$ 22.5 µl, 1 µM MgCl$_2$ 112.5 µl, ddH$_2$O add 45 mL) were inoculated with 5 mL of the culture and incubated for 1.5 h at 37 °C and 180 rpm. Cells were harvested by centrifugation for 5 min at 1800 × g and at RT and resuspended in 3 mL of supernatant. 1 µg of DNA was added to 100 µl of cell suspension. After 2 h of incubation at 37 °C and 900 rpm cells were plated on LB containing 5 mg/mL of Cm or 60 mg/mL Spc, depending on the transformed construct. Plates were incubated overnight at 37 °C. Single colonies were picked and transferred to a fresh LB plate containing appropriate antibiotics and incubated overnight at 37 °C. If a mutant already containing one of the selection markers was transformed, colonies were plated on both antibiotics separately for counter selection screening. Positive colonies were chosen for colony PCR. *Bacillus* colony material was cooked for 10 min in 20 µl of DMSO and used as a template. An overview of screening primers is provided in Table S4.

Expression and analysis of bacillus hybrids

To test for polyketide production, 5 mL YEME7 medium (4 g/L yeast extract, 10 g/L malt extract, 4 g/L glucose, 0.5 mL of 1 M NaOH per liter, add MOPS after autoclaving)[10a] with the appropriate antibiotics was inoculated with a single colony of *B. subtilis* strains. The culture was incubated in the dark at 37 °C and 250 rpm for 16-18 h. To check for bacillaene production or intermediates, the culture was mixed in the dark with methanol and shaken vigorously. The samples were spun down for 20 min at 10 °C and max speed and the supernatant subjected to UHPLC-HRMS analysis. Kinetex 2.6 µm XB-C18 100 A column (4.6 × 150 mm; Phenomenex) for 5 min at 10% acetonitrile (0.1% FA) in water (0.1% FA) followed by a gradient (10 to 100%) of acetonitrile (0.1% FA) in water (0.1% FA) over 11.5 min followed by 2 min 10% acetonitrile (0.1% FA) in water (0.1% FA) (1 mL/min). HPLC-ESI/MS spectra were obtained from a Q-Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer set to positive mode and coupled to a Dionex UltiMate™ 3000 UHPLC System (Thermo Scientific).

Computational details on the statistical coupling analysis

The sequences of BGCs annotated as *trans*-AT PKS in the antiSMASH database[14] and an in-house database of *trans*-AT PKS assembly lines were used in the statistical analysis. 2239 *trans*-AT PKS BGC deposited in the antiSMASH database[14] and an in-house database were extracted. An overview of the number of domain motifs is provided in Figure S5. Extraction of the tridomain sequences yielded 1194 amino acid sequences extracted from 516 clusters. From these sequences, the amino acid sequences indicated by the antiSMASH annotation of the domains, including 100 leading and trailing amino acids

in multidomain sequences, while taking gene termini into account. To prevent large highly gapped regions in MSAs of multidomain sequences, 15 amino acids following and leading consecutive domain annotations were extracted, instead of the full sequence linking the domains. The sequences obtained were then used to construct a multiple sequence alignment (MSA). We used several different algorithms to construct the MSAs. For the MSAs constructed with the MUSCLE algorithm[1], MUSCLE's standard settings were used. For MSAs constructed with MAFFT[15], the following options were used: '--auto --thread 5 --ep 0.0' or '--auto --thread 5 --ep 0.123'. MSAs with Clustal-Omega[16] constructed using the '--threads=5 --seqtype=Protein' options. The MSAs were then used for statistical coupling analysis (SCA) using Python scripts published by Rivoire et al.Rivoire, et al. [17] First, positions that contained more than 80% gaps and subsequently sequences containing more than 20% gaps were removed from the MSA and positions that contained more than 20% gaps were filtered from the alignment. Then, the sequences were weighed by the inverse of the total number of sequences in the MSA with which the sequence shares more than 80% sequence identity. The obtained filtered MSA and sequence weights were used to perform the statistical coupling analysis. Full details of SCA algorithm can be found elsewhere.[17] The final, filtered MSA contained 1190 sequences and 791 amino acid positions. The alignments that were used as input for the SCA were generated using MUSCLE 3.8.31, MAFFT or Clustal-Omega 1.2.4. For the MUSCLE alignments, the standard settings were used. For MAFFT alignments, the 'ep' offset value was set to 0.0 or 0.123, 5 threads were used and the algorithm was automatically selected using the --auto command. For Clustal-Omega alignments, 5 threads were used and seqtype was set to 'Protein'. The SCA was performed on a number of domain motifs with various alignments softwares to validate that the scission site suggested by the SCA is not an artefact of the alignment. The SCA matrices for the various domain motifs are given below. The EPIAII, HGTGT, NAHxVxE and TYPFx$_5$W motifs are indicated on the axes, indicating the *N*-terminus, active site and *C*-terminus of the KS domain and the *C*-terminus of the adapter domain, respectively. Since the NAHxVxE motif is not as highly conserved as many of the other motifs, the exact sequence of the motif in the consensus sequence varies slightly between the alignments (Figures S6-S13). In addition to the regular SCA, we also performed SCA on MSAs from which highly conserved (>80%) positions were filtered out after filtering for gapped sequences and positions. Because the highly conserved EPIAII, HGTGT and NAHxVxE motifs are highly conserved, these positions are not considered in these SCAs. As a result, they also do not show up in the SCA matrix. Therefore, the position in the filtered MSA is plotted on the axes labels. These positions can be related to the amino acids and positions using the color-coded consensus sequence of the conservation-filtered SCA below (Figure S14).

Supplementary Tables

Table S1. Strains used in this study.

Strain	Characteristics
E. coli DH5α (Invitrogen)	F– Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (rK–, mK+) *phoA supE44* λ– *thi*-1 *gyrA96 relA1*
E. coli BL21 Tuner™ (DE3, Novagen)	
S. plymuthica 4Rx13	NCBI accession number CP006250.1, NC_021591.1
S. plymuthica 4Rx13 Δ*oocR*	Δ^- *oocR*, Δ^+ *Kan^R*
S. plymuthica 4Rx13 Δ*oocQR*	Δ^- *oocQ, oocR*, Δ^+ *Kan^R*
B. subtilis DK1042[4]	B. subtilis 3610 comI^{Q12L}
B. subtilis DK1042 8PD9	Δ^- *pksM* (DH8-KS10)-*ymzB*, Δ^+ *Cm^R*, *comI*^{Q12L}
B. subtilis DK1042 8PD10 fusion site NAHVILEE	Δ^- *pksM* (DH8)-*ymzB*, Δ^+ *psyD* (dKS9-KS11), *SpcR*, *comI*^{Q12L}
B. subtilis DK1042 8PD10 fusion site LPTYPFx$_5$W	Δ^- *pksM* (DH8)-*ymzB*, Δ^+ *psyD* (dKS9-KS11), *SpcR*, *comI*^{Q12L}
B. subtilis DK1042 4PD10 fusion site NAHVILEE	Δ^- *pksM* (DH4)-*ymzB*, Δ^+ *psyD* (dKS9-KS11), *SpcR*, *comI*^{Q12L}
B. subtilis DK1042 4PD11 fusion site LPTYPFx$_5$W	Δ^- *pksM* (DH4)-*ymzB*, Δ^+ *psyD* (dKS10-KS11), *SpcR*, *comI*^{Q12L}

Table S2. Primers used for the expression construct and the *Serratia* deletion strains.

Name	Sequence
oocP_NheI_f	CATGCTAGCATGGAAAAGTATGACATTCAAAGTG
oocQ_NotI_r	CGCATGAATTCCTAACGGACGGACATTTTTAAC
oocP_noHis_F	AGCCAGGATCCGGAAAAGTATG
oocP_noHis_R	GCTGCTGCCCATGGTATATC
oocQ_noHis_F	AGCCAGGATCCGAGCGAG
KO_SOD_c22970_fw	GAGGACGCCTGTTATTTGCCACAACAGCTAAGCGAAGCCAGCGTCGATGTAAT TAACCCTCACTAAAGGGCGG
KO_SOD_c22970_rev	AACTCGTAAGCTCAGCCTTGGCGAGAAACGGCGTATAAGCGGCGCCTTCCTAA TACGACTCACTATAGGGCTC
KO_SOD_b01030_fwd	CCGGCAGACAAAGCCTATTACATCCTTAATGATGATCGGGCTTCGACGGAAAT TAACCCTCACTAAAGGGCGG

Table S3. Overview of Gibson assemblies and expression construct primers used.

Construct	Part	Template	Primer sequence
pBAD-oocQRC	pBAD	pBAD/Myc-His	TAGCGGCTTTGGATGTTTAGAATAGCGCCGTCGACCATCATCATCATCATC
			TTTATCAGGTACTCGCTCATCATGGTTAATTCCTCCTGTTAGCCCAAAAAAC
	$oocQ$ and $oocR_{KS011\text{-}DH\text{-}ACP\text{-}KS12}$	*S. plymuthica* 4Rx13	AACAGGAGGAATTAACCATGATGAGCGAGTACCTGATAAATTCCGGCGAG
			GCAGGTTCGGCGTTGCCGGCGATCCAGTATTGATCGGTCGCGAACGGATAAG
	$oocS_{ACP\text{-}C}$	*S. plymuthica* 4Rx13	CGACCGATCAATACTGGATCGCCGGCAACGCCGAACCTGCGCCTGTCGGC
			TGATGGTCGACGGCGCTATTCTAAACATCCAAAGCCGCTACCTCCTCCGGCGATAAACGATCGATC

Construct	Part	Template	Primer sequence
pBAD-oocQR	pBAD-$oocQR_{to\text{-}KS12}$	pBAD-oocQRC	CGATCAATACTGGATCAATAGCGCCGTCGACCATCATCATCATCATC
			GACGGCGCTATTGATCCAGTATTGATCGGTCGCGAACGGATAAG

Construct	Part	Template	Primer sequence
pBAD-oocQR-PsyDend	pBAD-$oocQR_{to\text{-}KS12}$	pBAD-oocQRC	TATGGCCTGGCCAAAGATAGAATAGCGCCGTCGACCATCATCATCATCATC
			GAGGGAATGAGCTCGGTAATGATCCAGTATTGATCGGTCGCGAACGGATAAG
	$psyD_{end}$ part 1	subcloned cluster #ref	CGACCGATCAATACTGGATCATTACCGAGCTCATTCCCTCGGGCATGCGC
			TCTCGTAAATCGGGGCGGACCAGCGGCTTCGCACGACCAGGCCACAGGAC
	$psyD_{end}$ part 2	subcloned cluster #ref	CTGGTCGTGCGAAGCCGCTGGTCCGCCCCGATTTACGAGAGCCGCGTTTG
			TGATGGTCGACGGCGCTATTCTATCTTTGGCCAGGCCATAACAAACGAATGATCTCATG

Construct	Part	Template	Primer sequence
pBAD-oocQR^{NAHVILEE}-PsyDend	$psyD$-pBAD-$oocQR_{to\text{-}KS12}$	pBAD-oocQR-PsyDend	TTTTGAAGCAACGGTTTTGGATTACCGAGCTCATTCCCTCGGGCATGCGCGATGAGCCGC
			TTGACGCCTTCCGGAGGCGCTTCCTCCAGCACGATATGTGCATTGGCGCCGCCAAAGCCG
	$psyD_{KS10}$ adapter domain	subcloned cluster #ref	CACATATCGTGCTGGAGGAAGCGCCTCCGGAAGGCGTCAAGGCACCCGGGGATGAGATGG
			GAGGGAATGAGCTCGGTAATCCAAAACCGTTGCTTCAAAAATGGGTAGGTAGGAAGTGTCAACGTTCGCCTGACTTG

Construct	Part	Template	Primer sequence
pBAD-oocQR-Lcn13-C	$oocS_{ACP\text{-}C}$-pBAD-$oocQR_{to\text{-}KS12}$	pBAD-oocQRC	CCCATGAACGTTACTGGATCGCCGGCAACGCCGAACCTGCGCCTGTCGGC
			GCCGGCGACGACTGCATGCCGATCCAGTATTGATCGGTCGCGAACGGATAAG
	$lcnB_{ACP\text{-}KS13}$	*G. sunshinyii* YC6258	CGACCGATCAATACTGGATCGGCATGCAGTCGTCGCCGGCAGCTACCCGG
			GCAGGTTCGGCGTTGCCGGCGATCCAGTAACGTTCATGGGAAAACGGATATACCGGCAGTTCCAG

Construct	Part	Template	Primer sequence
pBAD-oocQR-Lbm12-C	$oocS_{ACP\text{-}C}$-pBAD-$oocQR_{to\text{-}KS12}$	pBAD-oocQRC	TTGCCAAAAAACGTTGCTGGGCCGGCAACGCCGAACCTGCGCCTGTCGGCCAAATCATCGG
			GTATGGACCGCATCATTGAGGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGGCTGATGC
	$lbmD_{ACP\text{-}KS12}$	*G. sunshinyii* YC6258	CGACCGATCAATACTGGATCCTCAATGATGCGGTCCATACCGATGCGTCTGTGGTGCCGG
			GCAGGTTCGGCGTTGCCGGCCCAGCAACGTTTTTTGGCAAACGCATAGGTCGGCAGTGGC

Construct	Part	Template	Primer sequence
pBAD-oocQR^{NAHVILEE}-Lbm12-C	$lbmD_{ACP\text{-}KS12}$-$oocS_{ACP\text{-}C}$-pBAD-$oocQR_{to\text{-}KS12}$	pBAD-oocQR-Lbm12-C	TCCGTTTGCACCAGAGCGGCATTGGCTCAATGATGCGGTCCATACCGATGGTCCGGCCCGAACGCATCTTCAACATACTCCAGCACGATATGTGCATTGGCG
	$lbmD_{KS11}$ adapter domain	*G. sunshinyii* YC6258	CGCCAATGCACATATCGTGCTGGAGTATGTTGAAGATGCGTTCGGGCCGGAC
			CATCGGTATGGACCGCATCATTGAGCCAATGCCGCTCTGGTGCAAACGGA

Construct	Part	Template	Primer sequence
pBAD-oocQR-Lbm12^{NAHVILEE}-C	$lbmD_{ACP-KS12}$-$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQR-Lbm12-C	TCCGTTTGACGCGCGTTTATGCTGGGCCGGCAACGCCGAACCTGCG
			CCCCGTTTGAATCAGGCTCGGCGACCTCGGCCACGATCACATGTACATTC
	$lbmD_{KS12}$ adapter domain	G. sunshinyii YC6258	GAATGTACATGTGATCGTGGCCGAGGTCGCCGAGCCTGATTCAAACGGGG
			CGCAGGTTCGGCGTTGCCGGCCCAGCATAAACGCGCGTCAAACGGA

Construct	Part	Template	Primer sequence
pBAD-oocQR^{NAHVILEE}-Lbm12^{NAHVILEE}-C	$lbmD_{ACP-KS12}$-$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQR^{NAHVILEE}-Lbm12-C	TCCGTTTGACGCGCGTTTATGCTGGGCCGGCAACGCCGAACCTGCG
			CCCCGTTTGAATCAGGCTCGGCGACCTCGGCCACGATCACATGTACATTC
	$lbmD_{KS12}$ adapter domain	G. sunshinyii YC6258	GAATGTACATGTGATCGTGGCCGAGGTCGCCGAGCCTGATTCAAACGGGG
			CGCAGGTTCGGCGTTGCCGGCCCAGCATAAACGCGCGTCAAACGGA

Construct	Part	Template	Primer sequence
pBAD-oocQR-Pel9-C	$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	TTGCAAAAGAACGCCATTGGGCCGGCAACGCCGAACCTGCGCCTGTCGGCAAATCATCGG
			AAGCGATTTTTTTGCGAAGCGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGGCTGATGC
	$pelD_{ACP-KS9}$	metagenomic DNA of Mycale hentschelii	CGACCGATCAATACTGGATCGCTTCGCAAAAAAATCGCTTGCCGGATCCTTTGCCAACGA
			GCAGGTTCGGCGTTGCCGGCCCAATGGCGTTCTTTTGCAAAGGGATAGGTTGGCAAACTG

Construct	Part	Template	Primer sequence
pBAD-oocQR-Lbm11-C	$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	TTGCACCAGAGCGGCATTGGGCCGGCAACGCCGAACCTGCGCCTGTCGGCAAATC
			TCCAGCGTCATCCCGCCCTGGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGGC
	$lbmD_{ACP-KS11}$	G. sunshinyii YC6258	CGACCGATCAATACTGGATCCAGGGCGGGATGACGCTGGATGATACCACCAGGAC
			GCAGGTTCGGCGTTGCCGGCCCAATGCCGCTCTGGTGCAAACGGATACGTGGGTAAAC

Construct	Part	Template	Primer sequence
pBAD-oocQR-Lbm11mod-C	$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	TTGCACCAGAGCGGCATTGGGCCGGCAACGCCGAACCTGCGCCTGTCGGCAAATC
			TCAACACTGACGGCATCGCGGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGGC
	$lbmD_{DH-KR-ACP-KS11}$	G. sunshinyii YC6258	CGACCGATCAATACTGGATCCGCGATGCCGTCAGTGTTGATCAGACATCTTCTCC
			GCAGGTTCGGCGTTGCCGGCCCAATGCCGCTCTGGTGCAAACGGATACGTGGGTAAAC

Construct	Part	Template	Primer sequence
pBAD-oocQR-Pks5-C	$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	TCGCAAGAGATCGCTATTGGGCCGGCAACGCCGAACCTGCGCCTGTCGGCAAATC
			GCATCAATTTGCATGCCCCCGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGGC
	$pksL_{ACP-KS5}$	B. subtilis DK1042	CGACCGATCAATACTGGATCGGGGGCATGCAAATTGATGCGGAAACTGCAAGGAT
			GCAGGTTCGGCGTTGCCGGCCCAATAGCGATCTCTTGCGAAAGGATAGGCAGGTAAAC

Construct	Part	Template	Primer sequence
pBAD-oocQR-Pks5mod-C	$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	TCGCAAGAGATCGCTATTGGGCCGGCAACGCCGAACCTGCGCCTGTCGGCAAATC
			TTTTTCTCCGCTTTCGGCACGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGGC
	$pksL_{DH-KR-ACP-KS5}$	B. subtilis DK1042	CGACCGATCAATACTGGATCGTGCCGAAAGCGGAGAAAAAGACTGATCGTTCAAA
			GCAGGTTCGGCGTTGCCGGCCCAATAGCGATCTCTTGCGAAAGGATAGGCAGGTAAAC

Construct	Part	Template	Primer sequence
pBAD-oocQR-Lcn6mod-C	$oocS_{ACP-C}$-pBAD	pBAD-oocQRC	CGTGATCCGTACTGGGCCGGCAACGCCGAACCTGCGCCTG
			CAGGTACTCGCTCATCATGGTTAATTCCTCCTGTTAGCCCAAAAAACGGGTATGGAGAAAC
	$oocQ$ and $oocR_{KS011-DH-ACP-KS12}$	S. plymuthica 4Rx13	GAGGAATTAACCATGATGAGCGAGTACCTGATAAATTCCG
			GGTGGTATCCGGGGCGATCCAGTATTGATCGGTCGCGAAC
	$lcnA_{KR-ACP-KS6}$	S. plymuthica 4Rx13	GATCAATACTGGATCGCCCCGGATACCACCGCGGCAACCC
			TTCGGCGTTGCCGGCCCAGTACGGATCACGCGCAAATGGATAACCCG

166

Construct	Part	Template	Primer sequence
pBAD-oocQR-Gyn11-C	$oocS_{ACP-C}$-pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	CGGGAACGCTTTCCGGCCGGCAACGCCGAACCTGCGCCTG
			ACCGGCGCTTGCCAGGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGC
	$GynG_{KR-ACP-KS11}$	*G. sunshinyii* YC6258	GATCAATACTGGATCCTGGCAAGCGCCGGTCAGCCGGTGA
			TTCGGCGTTGCCGGCCGGAAAGCGTTCCCGTGCAAACGGATAGC

Construct	Part	Template	Primer sequence
pBAD-oocQR-PelEnd	pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	ACGATTTTAGACTTTAATAGCGCCGTCGACCATCATCATCATCATCATTGAGTTT
			GATCAGATAAATCCCGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGG
	$PelD_{KR-ACP-KS12-KR-ACP-KS013-ACP-TE}$	metagenomic DNA of *Mycale hentscheli*	GATCAATACTGGATCGGGATTTATCTGATCACAGGAGGGGCTGGAGGGCTTGGAT
			GTCGACGGCGCTATTAAAGTCTAAAATCGTCGTATAGACTTTAGGTTCGGACAGGAACAT

Construct	Part	Template	Primer sequence
pBAD-oocQR-OnnEnd	pBAD-$oocQR_{to-KS12}$	pBAD-oocQRC	ACTGTGTTGAGTAATAGCGCCGTCGACCATCATCATCATCATCATTGAGTTT
			CGGAAATCTTGATGATCCAGTATTGATCGGTCGCGAACGGATAAGTCGGCAGG
	$OnnJ_{KR-ACP-C-A-ACP-TE}$	metagenomic DNA of *Theonella swinhoei*	CAATACTGGATCATCAAGATTTCCGGCTCGCAGGTGCCGAACAGCAGTGTCC
			CGACGGCGCTATTACTCAACACAGTCAACTTCTCTTTGACGTGATTGGTGATCTGTTTGA

167

Table S4. Primers used for construction of Golden Gate constructs for the *Bacillus* mutants. This table was previously published in the thesis of Silke Reiter.

Name	Sequence	Templat	Reference for Template	(sub) cloned
Gent_*DraIII* Fw	ATCACACCGTGTTAGGTGGCGGTAC	TREX vector pIC20H-RL	JX668229.1[29]	pFus_A/*Spc*
Gent_*DraIII* Rv	ATC ACGTA GTGTAGGGATAACAGGGTAA			
GG_*Spe*ᴿ2 Fw	ttACTAGTTAACCATCGTGACGCGGC	pIC333	[28]	pGEM_GG_*Spc*ᴿ
GG_*Spe*ᴿ2 Rv	atCCTGCAGGCTAATTGAGAGAAGTTTC			
*Cm*ᴿ2.2Fus Fw	CTTACATAAGGAGGAACTACTATGGAGAAAAAAATCACTGGA	pACYC184	New England BioLabs (Ipswich, MA, USA)	pTopo_*Cm*ᴿ
GG_*Cm*ᴿ2.2 Rv	TGGTCTCATACGTTACGCCCCGCCC			
Phspac Fw	AGTTTAAACTACACAGCCCAGTCCAGACT	pMF37/pDGICZ	EU864235.1[32]	pTopo_GG_pH*Cm*ᴿ
pHspacFus Rv	TCCAGTGATTTTTTTCTCCATAGTAGTTCCTCCTTATGTAAG			
GG_pHspac Fw	AGGTCTCAACTCGTTTAAACTACACAGCCCAGT	pTopo_*Cm*ᴿ		
GG_*Cm*ᴿ2.2 Rv	TGGTCTCATACGTTACGCCCCGCCC			
GG_pH_*Cm*ᴿ3 Fw	AGGTCTCAAGctGTTTAAACTACACAGCCCAG	pTopo_GG_pH*Cm*ᴿ		pTOPO_pH*Cm*R3
GG_*Cm*ᴿ2.2 Rv	TGGTCTCATACGTTACGCCCCGCCC			
*Kan*ᴿ7 Fw	AGGTCTCACTAT *ACTAGT* CTGCGCTAGCATG	pCOLADuet-1	Novagen (Merck; Darmstadt, Germany)	pGEM_*Kan*ᴿ
*Kan*ᴿ7 Rv	AGGTCTCA GAGT TTAGAAAAACTCATCGAGCATCAAATG			
GG_Bs_*bae*KS8 Fw	AGGTCTCACTATGTTATACACCAGAGACAATTG	*B. subtilis* 3610 gDNA	DSMZ	pGEM_GG_*pks*MKS8
GG_Bs_*bae*KS8 Rv	TGG TCT CTA GC[TTCCTTCCAAAATAAGATGC			
GG_*psy*DKS9 Fw	AGGTCTCAAGctGTT CGT GAC ACG CCA GCT	pPSCG2	[12c]	pGEM_GG_*psy*DKS9
GG_*psy*DKS9 Rv	TGGTCTCTGAGTGATCAGAAGTCGCACTCAAGA			
*bae*KO_Bs Fw	AGGTCTCACTATCTTATGTCACGATTT	*B. subtilis* 3610 gDNA	DSMZ	pGEM_GG_*pks*KO
*bae*KO_Bs Rv	TGG TCT CTA GCTCTGCAAAATTTATATA			
GG_*ymzB* Fw	AGGTCTCACGTATGCACGATCTGTTACGA	*B. subtilis* 3610 gDNA	DSMZ	pJET_GG_*ymzB*
GG_*ymzB*E Rv	TGGTCTCTCGCCTTAATTAACATCAAAACTGAACC			
*PsyD*9 (1) Fw	atatatACTAGT GTTCGTGACACGCC	pPSCG2	[12c]	pGEM_*PsyD*9(1)
*PsyD*9 (1) Rv	t GTTTAAAC ATCCGGTTGAATGTAGCC			
*PsyD*9 (2) Fw	atatatACTAGT GTCGCATGTCACATTTG	pPSCG2	[12c]	pGEM_*PsyD*9(2)
*PsyD*9 (2) Rv	t GTTTAAAC CGCAAAGCCCTTCTC			
*PsyD*10 (1) Fw	atatatACTAGT GCGCCTCCGGA	pPSCG2	[12c]	pGEM_*PsyD*10
*PsyD*10 (2) Rv	t GTTTAAAC TAACAACAGATGTTGACGC			
*PsyD*11 (1) Fw	atatatACTAGT CACCGGTCGAGCA	pPSCG2	[12c]	pTopo_*PsyD*11
*PsyD*11 (2) Rv	t GTTTAAAC GCTCTATCTTTGGCCA			
pTOPO_*pksL*_KS4_up Fw	AGGTCTCACTATGAAGAAGCATCAGCAG	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_*pksL*_KS4_up
pTOPO_*pksL*_KS4_up Rv	TGGTCTCTagcTTCCTCAAGGATAATATGTG			
pTOPO_*baeL*_KS4_down Fw	AGGTCTCAAgctTTTATCCCTGAAGAATCAC	*B. amyloliquefaciens* gDNA	DSMZ	pTOPO_*baeL*_KS4_down
pTOPO_*baeL*_KS4_down Rv	TGGTCTCTGAGTCGGCGTGACTCT			
pTOPO_*baeL*_KS5_up Fw	AGGTCTCACTATTATCAGCTTTACGGAGC	*B. amyloliquefaciens* gDNA	DSMZ	pTOPO_*baeL*_KS5_up
pTOPO_*baeL*_KS5_up Rv	TGGTCTCTagcTTCTTCGATGACGATAT			
pTOPO_*pksL*_KS5_down Fw	AGGTCTCAAgctTATGCTCCGGAACC	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_*pksL*_KS5_down
pTOPO_*pksL*_KS5_down Rv	TGGTCTCTGAGTACTCGTGTGCTAAGTA			
pTOPO_*pksM*_up Fw	AGGTCTCACGTAAAGAGAGGAGTGGGA	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_*pksM*_up
pTOPO_*pksM*_up Rv	TGGTCTCTCGCCTTAATTAACGCAAATACATT			
*mmpD*KS2down Fw	AGGTCTCAAgct CAGCCACCCCT	pFusA_*mmpD*_KS2-3		pTOPO_*mmpD*_KS2_down
*mmpD*KS2down Rv	TGGTCTCTGAGT GCCAGATCCAAGG			
*mmpD*KS3up Fw	AGGTCTCACTAT CATGTACCAGGAGT	pFusA_*mmpD*_KS2-3		pTOPO_*mmpD*_KS3_up
*mmpD*KS3up Rv	TGGTCTCTagcT TCCTCGATGAT			
*pksL*KS5down_d Fw	AGGTCTCACGTATATGCTCCGGAACCTGTGG	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_*pksL*_KS5_down2
*pksL*KS5down_d Rv	TGGTCTCTCGCCTTAATTAAACTCGTGTGCTAAGTA			
pTOPO_*baeL*_KS4-5 Fw	atatatACTAGTTATGCTCCGGAACCTGTGGA	*B. amyloliquefaciens* gDNA	DSMZ	pTOPO_*baeL*KS4-5
pTOPO_*baeL*_KS4-5 Rv	tGTTTAAACTGATGCTTGCTTGCACGTCC			
pTOPO_*pksL*_down1 Fw	atatatACTAGTTATGCTCCGGAACCTGTGGA	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_*pksL*_down1
pTOPO_*pksL*_down1 Rv	tGTTTAAACTGATGCTTGCTTGCACGTCC			
pTOPO_*pksL*_down2 Fw	atatatACTAGTCATGTGTGAGCTCCTTCGG	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_*pksL*_down2
pTOPO_*pksL*_down2 Rv	tGTTTAAACCCCACTCCTCTCTTATTTGAAAGT			
pksL Fw	AGGTCTCACTATATATAGGAGGCCG	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_GG_*pksL*_hom
pksL Rv	TGGTCTCTGAGTTTATTTGAAAGTTTTCC			
pksM Fw	AGGTCTCACGTAAGAGAGGAGTGG	*B. subtilis* 3610 gDNA	DSMZ	pTOPO_GG_*pksM*_hom
pksM Rv	TGGTCTCTCGCCCGCAAATACATT			
Screening primer name	Sequence	Primer binding		Type of hybrid
CP *Cat*2.2 Fw	CGTGGCCAATATGGACAACTTC	Inside *Cm*ᴿ marker		Internal/terminal
CP2_*Spc* Fw	AAGTGGGAAGGACTATATTCAAAGG	Inside SpcR marker		Internal/terminal
CP all Bs Rv	CATCCCGATGGACAAACTTGG	Downstream of "*ymzB*"-DGH		Terminal
PKS_L/M Check Rv	AACAGCAGTGTCGGAGCAAG	Downstream of "*pksM*"-DGH		Internal

Name	Sequence	Templat	Reference for Template	(sub) cloned
OnnJ I Fw	atatatACTAGTTACCTTCCCCAAAGGGC	pTSTA11	[12a]	pTopo_*onnJ*I
OnnJ I Rv	tGTTTAAACTTGCTCAAACCGGTAATTTC			
OnnJ II Fw	atatatACTAGTCTATTACACCGAAGACCG	pTSTA11	[12a]	pTopo_*onnJ*II
OnnJ II Rv	tGTTTAAACCTGATCCAGGCCGAATA			
OnnJ III Fw	atatatACTAGTATCCCAACATTCTGGGTCA	pTSTA11	[12a]	pTopo_*onnJ*III
OnnJ III Rv	tGTTTAAACCGAGACCGATGCGTCGT			
OnnJ IV Fw	atatatACTAGTACAACAGAGTCTCGAAGA	pTSTA11	[12a]	pTopo_*onnJ*IV
OnnJ IV Rv	tGTTTAAACCTAACTCAACACAGTCAACT			
GG_DifLKS14 Fw	AGGTCTCAAgctGCGCCGCAAACCGTG	gDNA *B. amyloliquefaciens* FZB42	DSMZ	pTOPO_GG_*difL*
GG_DifLKS14 Rv	TGGTCTCTGAGTCCGCAAGCTCTTCAAGC			
DifL I Fw	atatatACTAGTGCGCCGCAAACCGTGC	gDNA *B. amyloliquefaciens* FZB42	DSMZ	pTopo_*difL*I
DifL I Rv	tGTTTAAACTTACGGTCTTGTTTTGGCGG			
GG_pksXKS4 Fw	AGGTCTCACTATTTAGTCGGAGCTGAAGAA	gDNA *B. subtilis* 3610	DSMZ	pGEM_GG_*pksM*KS4
GG_pksXKS4 Rv	TGGTCTCTA GCTTCCTCAAGGATAATATGTGC			
GG_OnnJKS11 Fw	AGGTCTCAAgctTACCTTCCCCAAAGGGCGC	pTSTA11	[12a]	pGEM_GG_*onnJ*KS11
GG_OnnJKS11 Rv	TGGTCTCTGAGTGCCGTAACGCATCAAGATCC			
GG_PsyDKS11 Fw	AGGTCTCAAgctGCACCGGTCGAGCAAAC	pPSCG2	[12c]	pGEM_GG_*psyD*KS11
GG_PsyDKS11 Rv	TGGTCTCTGAGTGTAACAACAGATGTTGACGCG			

Table S5. Golden Gate assemblies in *Bacillus* hybrids.

Construct	Part	Template plasmid
pFusA_PRBM8PD9 Golden Gate	backbone	pFus_A/Spc
	upstream genome homology	pGEM_GG_*pksM*KS8
	first gene fragment	pGEM_GG_*psyD*KS9
	resistance cassette	pTopo_GG_pHCm^R
	downstream genome homology	pJET_GG_*ymzB*

Construct	Part	Template plasmid
PRB4PD9 Golden Gate	backbone	pFus_A/Gm
	upstream genome homology	pGEM_GG_pksMKS4
	first gene fragment	pGEM_GG_psyDKS11
	resistance cassette	pTopo_GG_SpcR
	downstream genome homology	pJET_GG_*ymzB*

Construct	Part	Template plasmid
PRB8OJ11 Golden Gate	backbone	pFus_A/Spc
	upstream genome homology	pGEM_GG_pksMKS8
	first gene fragment	pGEM_GG_onnJKS11
	resistance cassette	pTopo_GG_pHCm^R
	downstream genome homology	pJET_GG_*ymzB*

Construct	Part	Template plasmid
PRB8DL14 Golden Gate	backbone	pFus_A/Gm
	upstream genome homology	pGEM_GG_pksMKS8
	first gene fragment	pTOPO_GG_difL
	resistance cassette	pTopo_GG_SpcR
	downstream genome homology	pJET_GG_*ymzB*

Construct	Part	Template plasmid
pFusA/S/C_Kan^R Golden Gate	backbone	pFus_A/Spc
	spacer	pGEM_GG_Kan^R
	resistance cassette	pTopo_GG_pHCm^R
	downstream genome homology	pJET_GG_*ymzB*

Construct	Part	Template plasmid
pFusA/G/S_Kan^R Golden Gate	backbone	pFus_A/Gm
	spacer	pGEM_GG_Kan^R
	resistance cassette	pGEM_GG_Spc^R

	downstream genome homology	pJET_GG_*ymzB*	

Construct	Part	Template plasmid
pFusA/S_*pks*_KO Golden Gate	backbone	pFus_A/Spc
	upstream genome homology	pGEM_GG_pksKO
	Cm^R resistance cassette	pTopo_GG_pHCm^R
	downstream genome homology	pJET_GG_*ymzB*

Construct	Part	Template plasmid
pFusA/S_*pks*_TEKO Golden Gate	backbone	pFus_A/Spc
	upstream genome homology	pTOPO_GG_TE_KO
	Cm^R resistance cassette	pTopo_GG_pHCm^R
	downstream genome homology	pJET_GG_*ymzB*

	Construct	Acceptor plasmid	Insert plasmid
Restriction cloning	pFusA_PD9(1)	pFusA/G/S_Kan^R	pGEM_*psyD*9(1)
	pFusA_PD9(2)	pFusA/S/C_Kan^R	pGEM_*psyD*9(1)
	pFusA_PD10	pFusA/G/S_Kan^R	pGEM_*psyD*10
	pFusA_PD11	pFusA/S/C_Kan^R	pGEM_*psyD*11
	pFus_A/G/C_psyD11	pFus_A/S/C_Kan^R	pTopo_psyD11
	pFus_A/G/S_onnJI	pFusA/G/S_Kan^R	pTopo_onnJI
	pFus_A/S/C_onnJII	pFus_A/S/C_Kan^R	pTopo_onnJII
	pFus_A/G/S_onnJIII	pFusA/G/S_Kan^R	pTopo_onnJIII
	pFus_A/S/C_onnJIV	pFus_A/S/C_Kan^R	pTopo_onnJIV
	pFUS_A/S/C_difLI	pFus_A/S/C_Kan^R	pTopo_difLI

Construct	Part	Template	Primer sequence
pBAD-4PD11 Gibson cloning	pBAD	pBAD/Myc-His	GTAAATGCTTGATGAACAAGTGATCTCAATAGCGCCGTCGACCATCATC CTTTGGCGTATAGCCCGCATCCATGGTTAATTCCTCCTGTTAGCC
	upstream genome homology	*B. subtilis* DK1042	GGCTAACAGGAGGAATTAACCATGGATGCGGGCTATACGCCAAAG TTCCAAAACGGGAATCCAATAACGCACCCTCTCAAATGGG
	*psyD*DH-ACP-TE	subcloned cluster #ref	AGGGTGCGTTATTGGATTCCCGTTTTGGAAGAGAAGGCGG GCGTCACGATGGTTAGCTCTATCTTTGGCCAGGCCATAAC
	spcR resistance cassette	pIC333 #ref	GGCCAAAGATAGAGCTAACCATCGTGACGCGGCATTCTAG ATCCAGCACAGCTGACTAATTGAGAGAAGTTTCTATAGAATTTTTCATAT ACTTAACGAG
	downstream genome homology	*B. subtilis* DK1042	ACTTCTCTCAATTAGTCAGCTGTGCTGGATATCAATTGTATATAC GATGATGGTCGACGGCGCTATTGAGATCACTTGTTCATCAAGCATTTAC

Construct	Part	Template	Primer sequence
pBAD-4PD10-NAHVILEE	backbone	pBAD-4PD11	AGCACCGGTCGAGCAATTCCCGTTTTGGAAGAGAAGGCGG GCCTTCCGGAGGCGCTTCCTCAAGGATAATATGTGCGTTTGAAC
Gibson cloning	*psyD*ACP-KS11	subcloned *psy* cluster	ATTATCCTTGAGGAAGCGCCTCCGGAAGGCGTCAAGGCAC TTCCAAAACGGGAATTGCTCGACCGGTGCTTCCCCCAGAATG

Construct	Part	Template	Primer sequence
pBAD-4PD10-LPTYPFx$_5$W	backbone	pBAD-4PD11	AGCACCGGTCGAGCAATTCCCGTTTTGGAAGAGAAGGCGG
			AATGAGCTCGGTAATCCAATAACGCACCCTCTCAAATGGG
	$psyD_{ACP-KS11}$	subcloned psy cluster	AGGGTGCGTTATTGGATTACCGAGCTCATTCCCTCGGGCATG
Gibson cloning			TTCCAAAACGGGAATTGCTCGACCGGTGCTTCCCCCAGAA

Construct	Part	Template	Primer sequence
pBAD-8PD10-NAHVILEE	backbone	pBAD-4PD11	CTTATTTTGGAAGAAGCGCCTCCGGAAGGCGTCAAGGCAC
			AGCAATTGTCTCTGGCATGGTTAATTCCTCCTGTTAGCCCAAAAAACGGG
	$psyD_{ACP-KS11}$	subcloned psy cluster	GAGGAATTAACCATGCCAGAGACAATTGCTTCACCGCAGG
Gibson cloning			GCCTTCCGGAGGCGCTTCTTCCAAAATAAGATGCGCATTTGATC

Construct	Part	Template	Primer sequence
pBAD-8PD10-LPTYPFx$_5$W	backbone	pBAD-4PD11	AGGGAGCGCTACTGGATTACCGAGCTCATTCCCTCGGGCATG
			AGCAATTGTCTCTGGCATGGTTAATTCCTCCTGTTAGCCCAAAAAAC
	$psyD_{ACP-KS11}$	subcloned psy cluster	GAGGAATTAACCATGCCAGAGACAATTGCTTCACCGCAGG
Gibson cloning			AATGAGCTCGGTAATCCAGTAGCGCTCCCTTGCAAACGGATAG

Table S6. BlastP hits for OocQ. Perc. Ident.: percentage identity, Acc. Len.: Accession length. Query issued on 26.11.2021.

Description	Scientific Name	Query Cover	E value	Per. Ident	Acc. Len	Accession
hypothetical protein [Serratia plymuthica]	*Serratia plymuthica*	1	3E-82	1	119	WP_004944605.1
hypothetical protein [Serratia plymuthica]	*Serratia plymuthica*	1	2E-81	0.9916	119	WP_062790739.1
hypothetical protein [Serratia marcescens]	*Serratia marcescens*	1	1E-75	0.916	120	WP_074054222.1
hypothetical protein [Serratia marcescens]	*Serratia marcescens*	1	9E-72	0.8655	119	WP_073528974.1
hypothetical protein [Dickeya chrysanthemi]	*Dickeya chrysanthemi*	0.99	2E-67	0.822	120	WP_226052029.1
hypothetical protein [Dickeya sp. NCPPB 3274]	*Dickeya* sp. NCPPB 3274	0.99	3E-67	0.822	119	WP_042861989.1
hypothetical protein [Dickeya chrysanthemi]	*Dickeya chrysanthemi*	0.99	3E-67	0.822	119	WP_040001540.1
hypothetical protein [Dickeya]	*Dickeya*	0.99	1E-66	0.822	119	WP_033576333.1
hypothetical protein [Dickeya solani]	*Dickeya solani*	0.99	4E-66	0.8136	120	WP_223849435.1
hypothetical protein [Dickeya solani]	*Dickeya solani*	0.99	5E-66	0.8136	119	WP_022635046.1
hypothetical protein [Dickeya dianthicola]	*Dickeya dianthicola*	0.99	7E-66	0.8136	121	WP_223303756.1
hypothetical protein [Dickeya dianthicola]	*Dickeya dianthicola*	0.99	7E-66	0.8136	120	WP_024106520.1
hypothetical protein [Musicola paradisiaca]	*Musicola paradisiaca*	0.99	1E-65	0.7966	119	WP_012765047.1
hypothetical protein [Dickeya oryzae]	*Dickeya oryzae*	0.99	8E-65	0.7966	120	WP_210174681.1
hypothetical protein [Dickeya oryzae]	*Dickeya oryzae*	0.99	9E-65	0.7966	121	WP_226093247.1
hypothetical protein [Dickeya oryzae]	*Dickeya oryzae*	0.99	3E-64	0.7966	120	WP_153508288.1
hypothetical protein [Dickeya oryzae]	*Dickeya oryzae*	0.99	3E-63	0.7881	120	WP_210194656.1
hypothetical protein [Dickeya zeae]	*Dickeya zeae*	0.99	4E-63	0.7797	120	WP_016940740.1
hypothetical protein [Aquimarina sp. RZ0]	*Aquimarina* sp. RZ0	0.92	2E-09	0.3158	129	WP_149624478.1

Table S7. Overview of the foreign PKS parts used in this study.

Name	PKS source and compound	Domain sequence
BasC	*bas* PKS, basiliskamide	C
PsyD$_{term}$	*psy* PKS, psymberin	ACP-KS-?-DH-ACP-TE
LbmD12	*lbm* PKS, lobatamide	ACP-KS
LbmD11		ACP-KS
LbmD11-rec		DH-KR-ACP-KS
GynG11	*gyn* PKS, gynuellalide	KR-ACP-KS
LcnA6	*lcn* PKS, lacunalide	KR-ACP-KS
LcnB13		ACP-KS
PksL5	*pks* PKS, bacillaene	ACP-KS
PksL5-rec		DH-KR-ACP-KS
PelD9	*pel* PKS, peloruside	ACP-KS
PelD11	*pel* PKS, peloruside	KR-ACP-KS-KR-ACP-KS-TE
Onnamide	*onn* PKS, onnamide	KS-KR-ACP-C-A-PCP-TE

Table S8. Overview of consensus retention times of **1** and **2-5cyc** (see Fig. S17), as extracted from the combined UHPLC-MS traces. n.d. indicates not determined.

Compound	Consensus retention time (min)
1	6.27
2cyc	7.82
2cyc$_a$	7.82
2cyc$_b$	6.6
3	6.16
3cyc$_a$	6.16
3cyc$_b$	5.87
4	6.64
4cyc$_a$	6.64
4cyc$_b$	5.89
5	7.54
5cyc$_a$	n.d.
5cyc$_b$	6.46
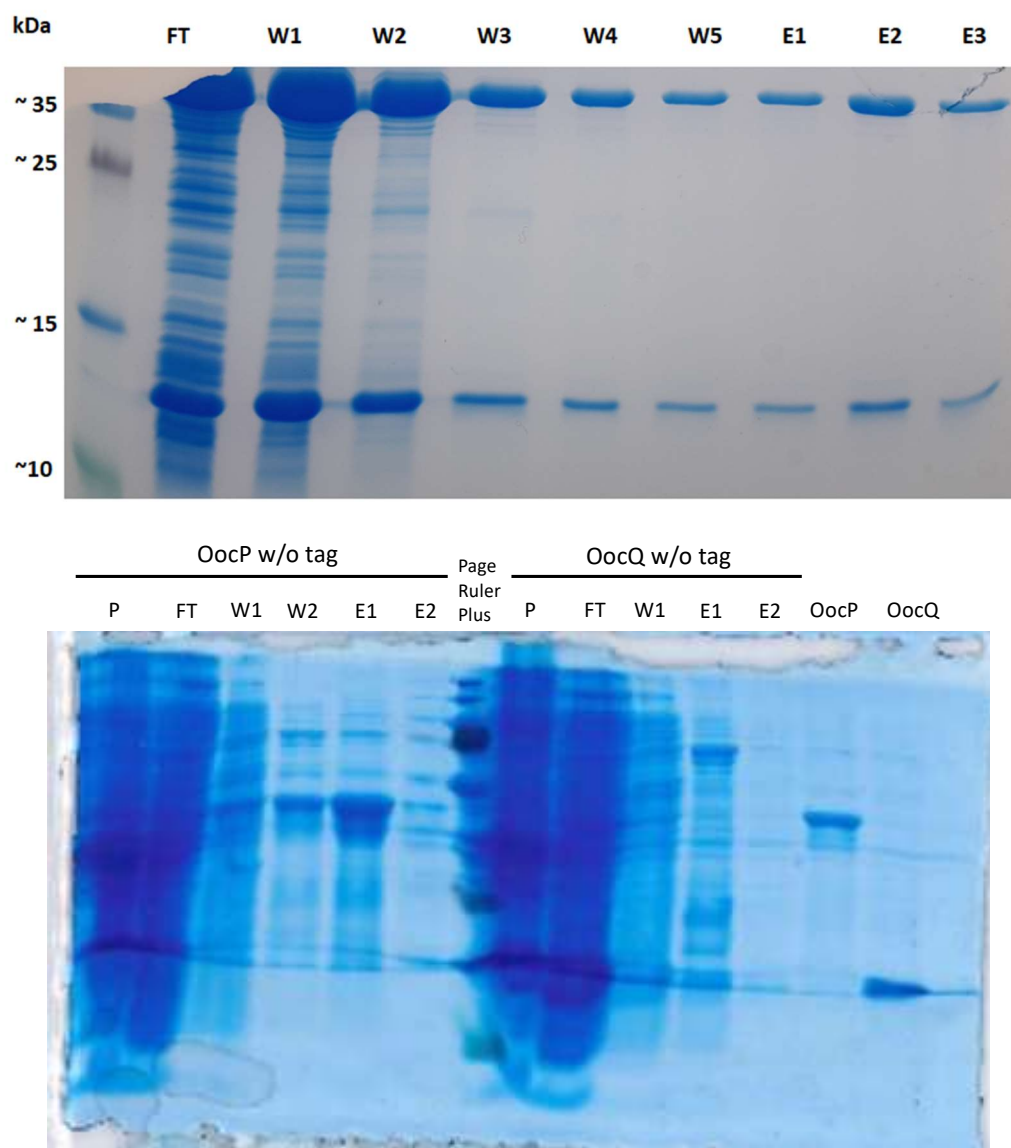
Figure S1. Coomassie-stained 12% SDS-PAGE gels of OocP and OocQ (expected molecular mass: 45.5 kDa and 14.0 kDa, respectively) post Ni-NTA purification. Top: co-expression of native OocP and OocQ, bottom: separate expression. As controls, His-tagged OocP (lane labelled OocP) and OocQ (lane labelled OocQ) were purified. P: pellet, FT: flowthrough, W: wash step, E: elution.
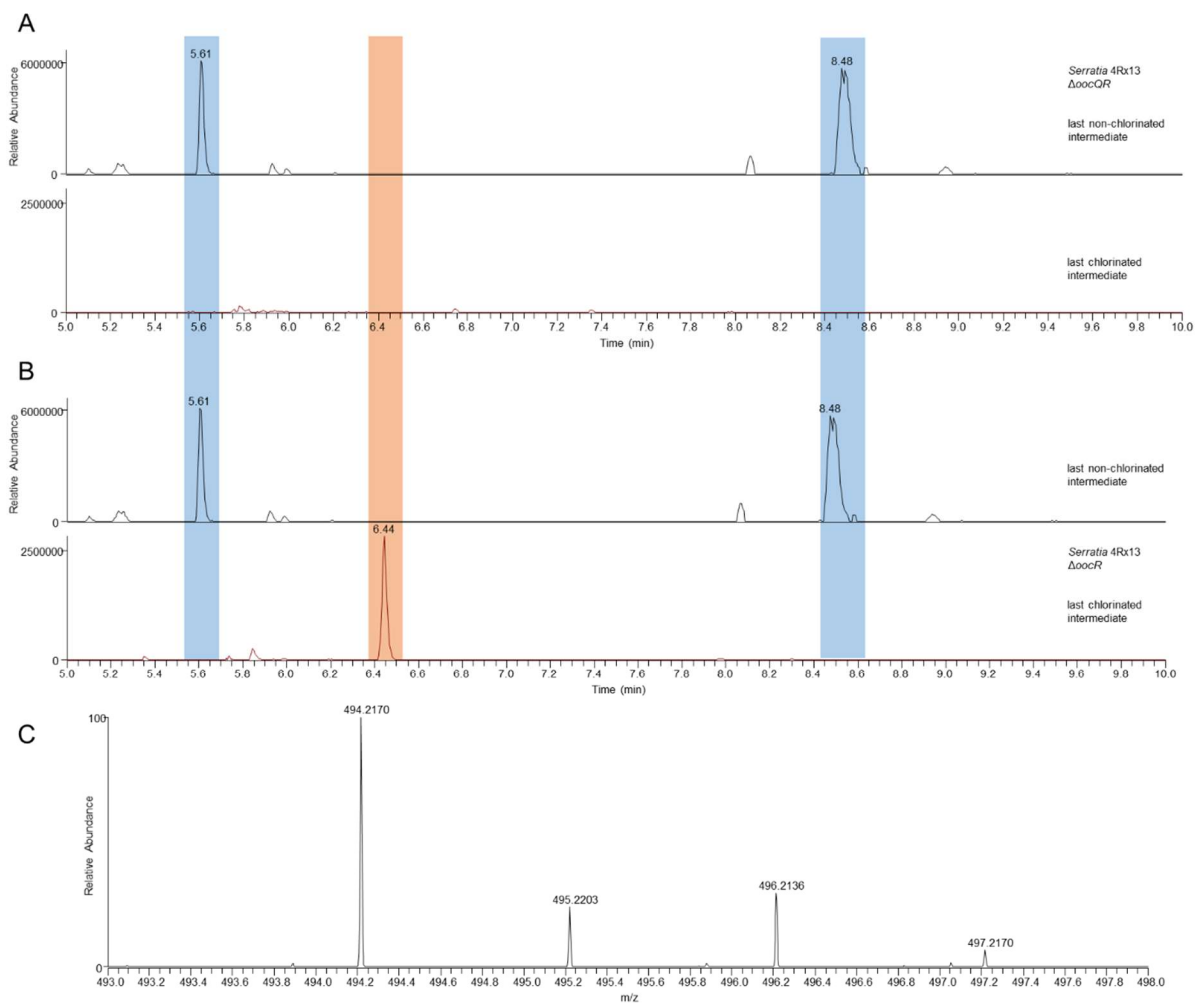
Figure S2. Intermediate analysis of *Serratia plymuthica* 4Rx13 Δ*oocQR* and *Serratia plymuthica* 4Rx13 Δ*oocR*. Extracted ion chromatograms for the last non-chlorinated intermediate ($[C_{21}H_{32}O_9+NH_4]^+$ as 446.2385) and the first chlorinated intermediate ($[C_{22}H_{33}ClO_9+NH_4]^+$ as 494.2151) according to the biosynthetic model. A) Data for *Serratia plymuthica* 4Rx13 Δ*oocQR*. B) Data for *Serratia plymuthica* 4Rx13 Δ*oocR*. C) Mass spectrum of the chlorinated intermediate in *Serratia plymuthica* 4Rx13 Δ*oocR* with the typical chlorination isotope pattern.

MEKYDIQSETQRKAFDLMP**H**FFLDQEEQANF**H**FMM**H**MRLLLNAPEFMAT
FERDLFEKKLADLQAKCPDLANMDCADTFIKMKSYDFSNMDR**H**TFQ**H**MIN
DASNPPIIAKGFLNDTKAVQQWT**H**EYLIE**H**YKDTEIIAVGYDEKETSLKKLK
LEKILRSQLDKDSKVSYYINNSAEIFNDYPDLIDEVGAEKILDLFYG**H**SANSF
SQLFVGNLRTWGTNW**H**QGNDISCALMISGVKRWYFVDPRLGYILRPFFD
GANGMSAKMDARLDMNF**H**KI**H**SPLYAYAPKFYVDLEPGDVIFFTKYWP**H**
AVINTTPLQIMANMRMTEVNLDTMTKGKDVPTLMPVYDNILNSDPSFIKFK
FDIFNNLGKKSKTIGDENYFSAYTSTSDVLTNGDKQ

Figure S3. Amino acid sequence of OocP with all His residues highlighted in pink.

Figure S4. WebLogo representation of manual alignment of extended KS sequences.
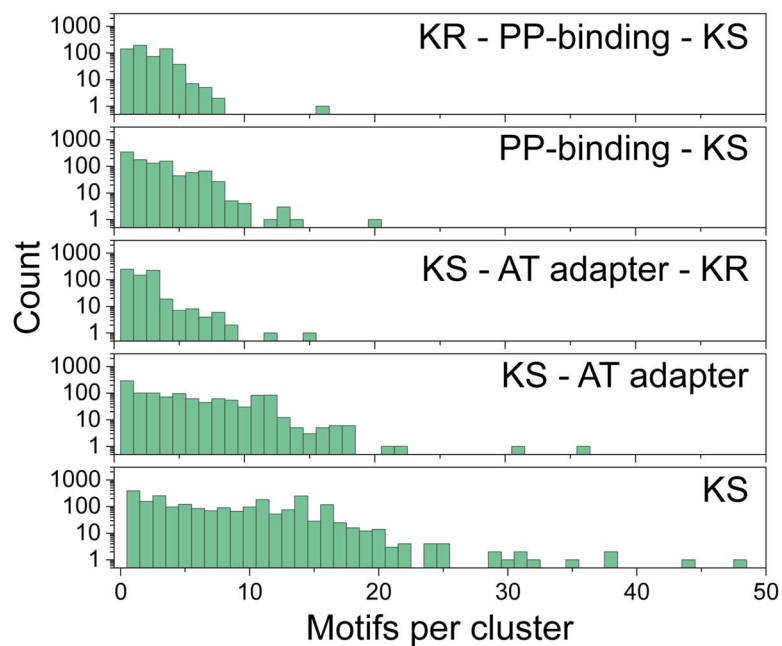
Figure S5. Histograms of the number of domain motifs found in all entries deposited in the antiSMASH database as of June 6th, 2021.
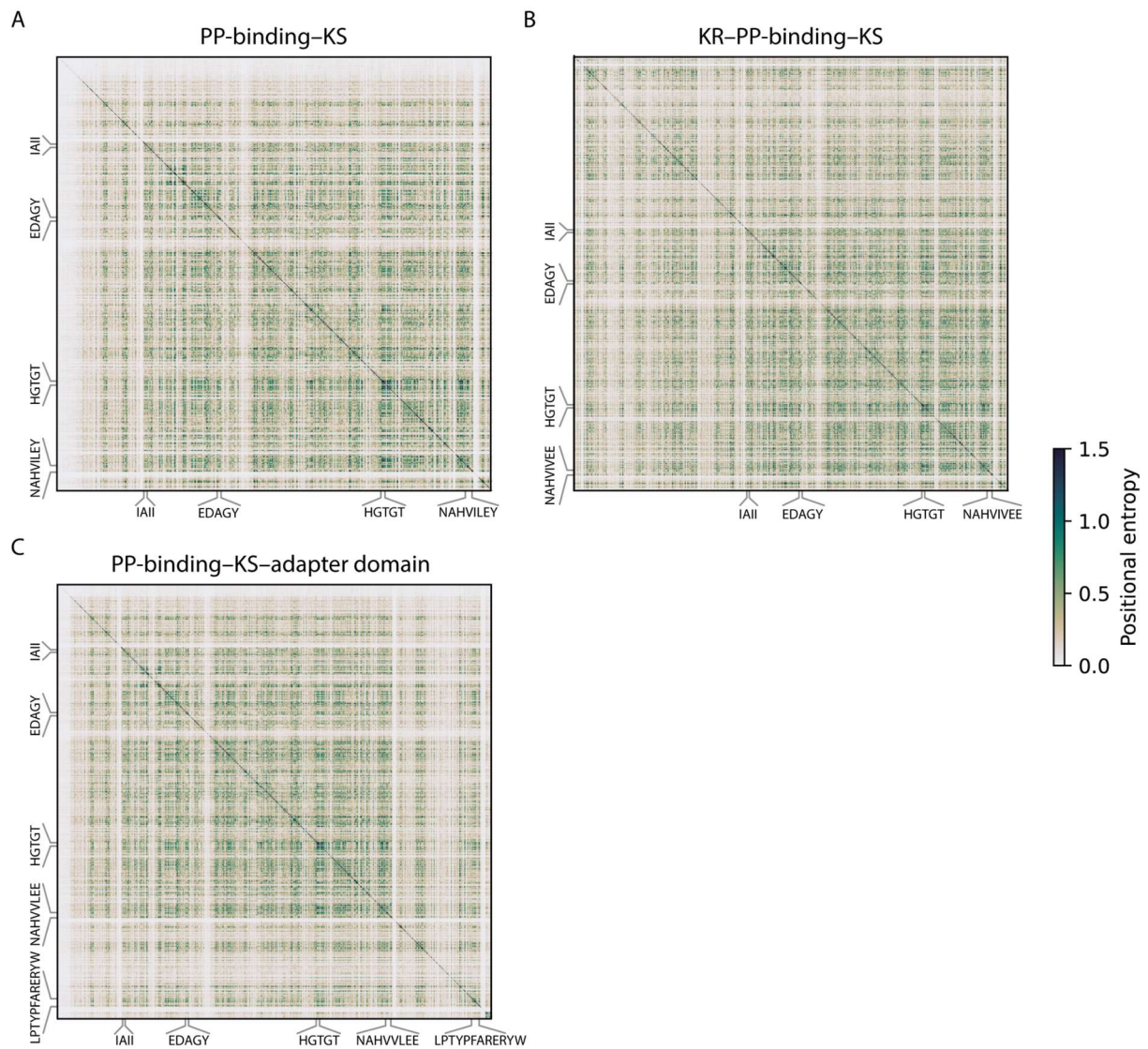
Figure S6. SCA matrices of alignments obtained using the MUSCLE algorithm of A) PP-binding–KS, B) KR–PP-binding–KS and C) PP-binding–KS–adapter domain. The IAII, EDAGY, HGTGT, and the consensus sequence of the NAHVILEE and LPTYPFx₅W motifs obtained from each MSA are indicated on the axes, indicating the N-terminus, active site and *C*-terminus of the KS domain and the *C*-terminus of the adapter domain, respectively.

181
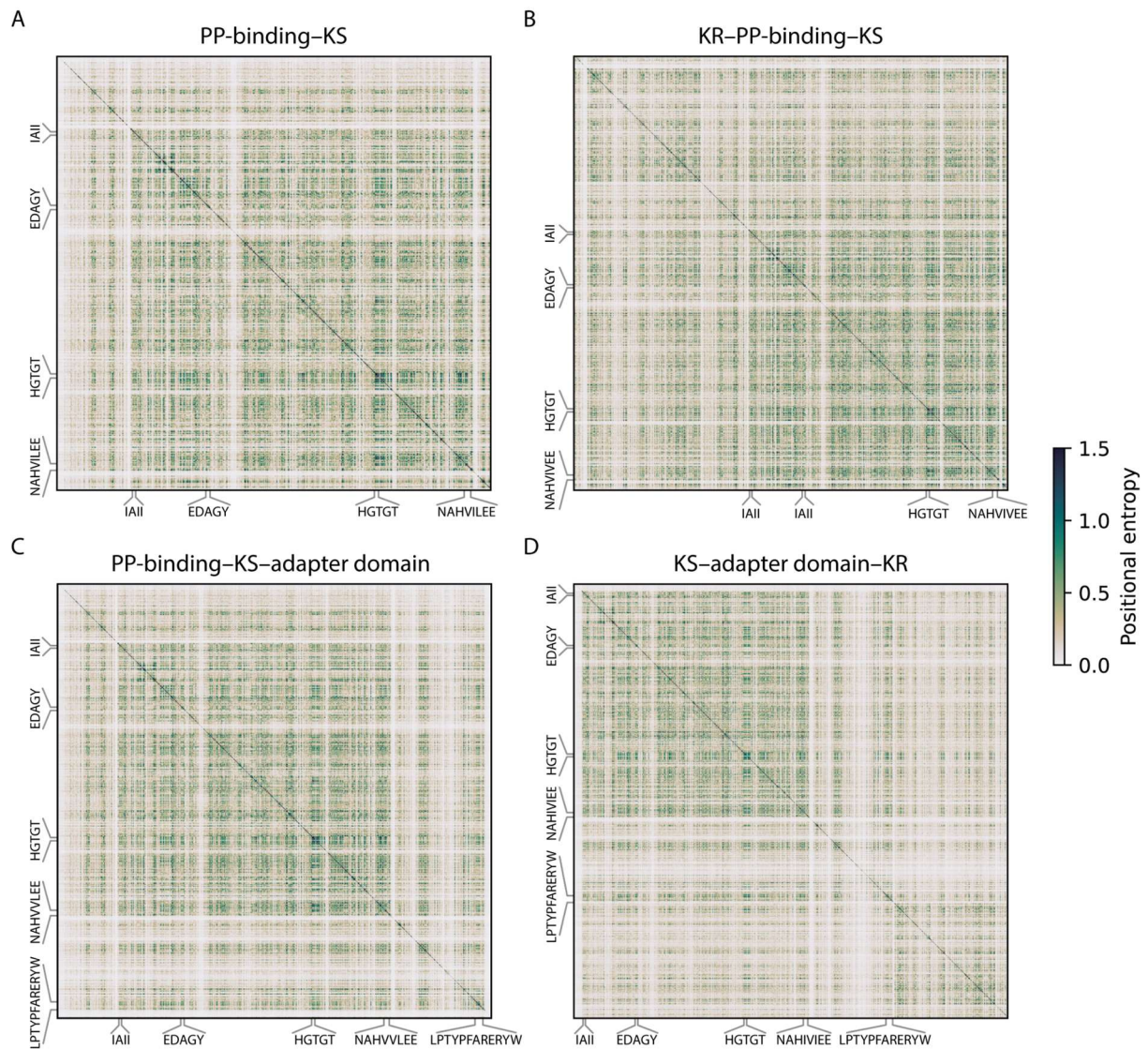
Figure S7. SCA matrices of alignments obtained using the Clustal-Omega algorithm of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR. The IAII, EDAGY, HGTGT, and the consensus sequence of the NAHVILEE and LPTYPFx$_5$W motifs obtained from each MSA are indicated on the axes, indicating the N-terminus, active site and *C*-terminus of the KS domain and the *C*-terminus of the adapter domain, respectively.
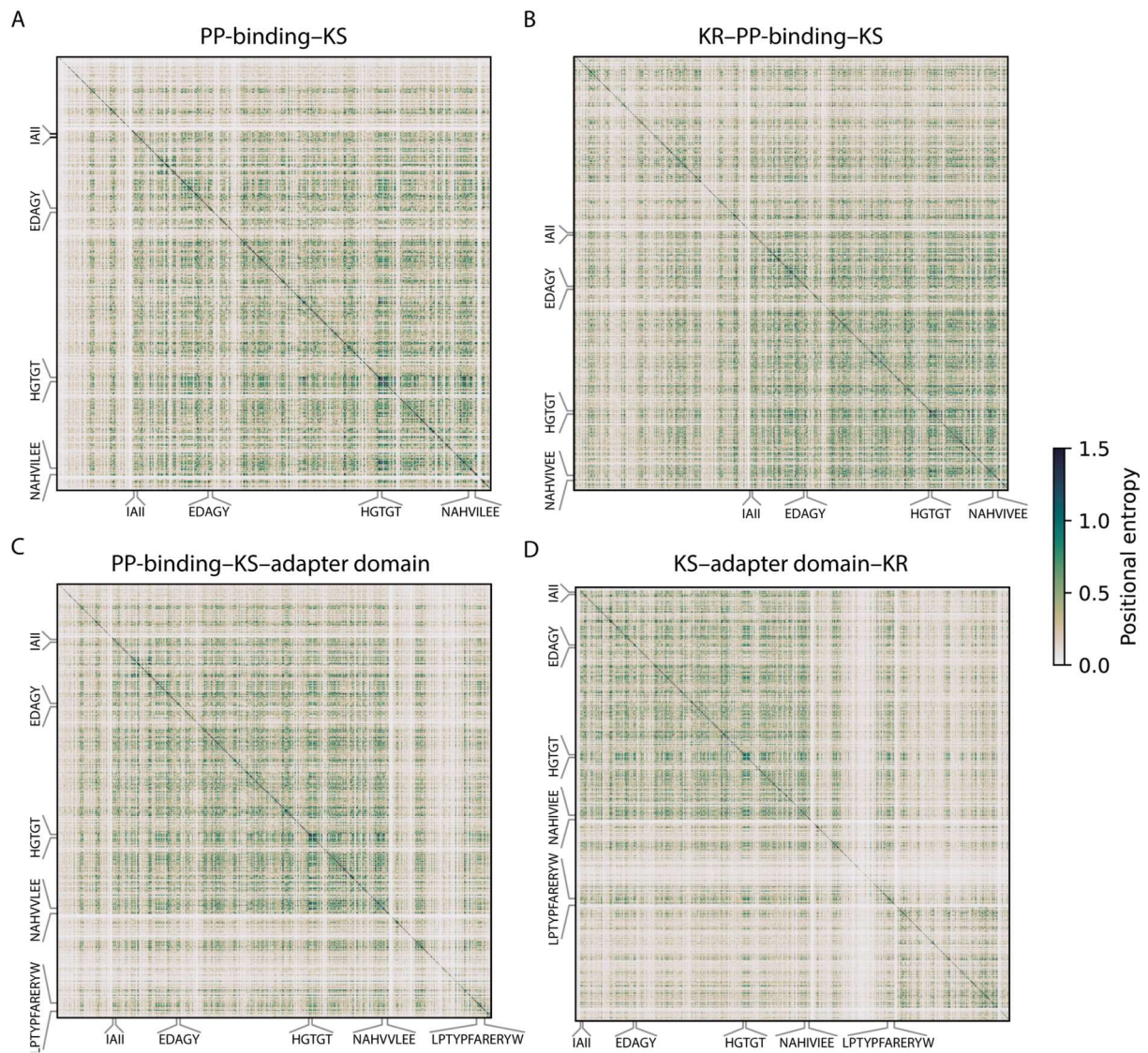
Figure S8. SCA matrices of alignments obtained using the MAFFT algorithm and an ep offset of 0.0 of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR. The IAII, EDAGY, HGTGT, and the consensus sequence of the NAHVILEE and LPTYPFx$_5$W motifs obtained from each MSA are indicated on the axes, indicating the N-terminus, active site and *C*-terminus of the KS domain and the *C*-terminus of the adapter domain, respectively.
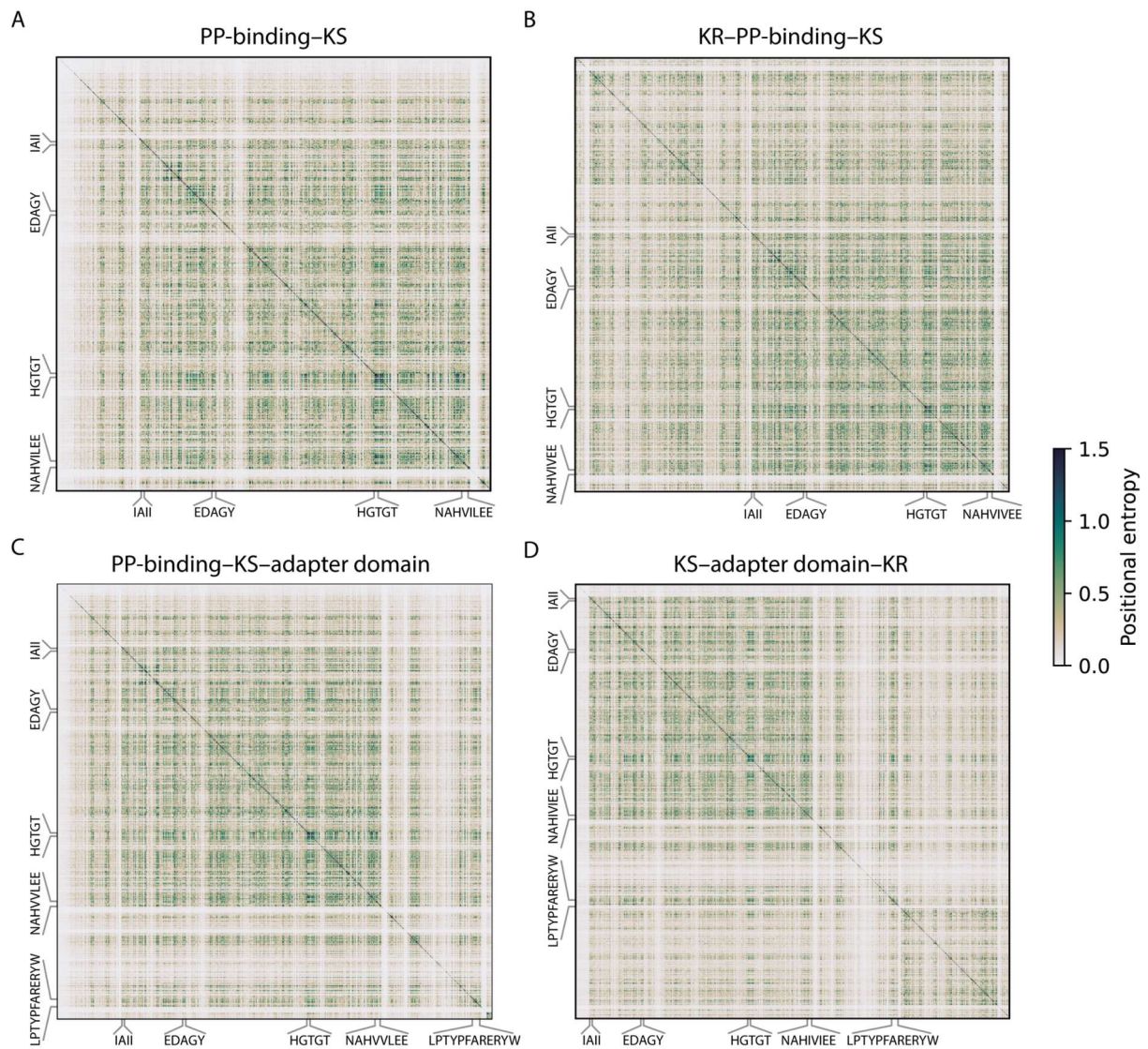
Figure S9. SCA matrices of alignments obtained using the MAFFT algorithm and an ep offset of 0.123 of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR. The IAII, EDAGY, HGTGT, and the consensus sequence of the NAHVILEE and LPTYPFx$_5$W motifs obtained from each MSA are indicated on the axes, indicating the N-terminus, active site and *C*-terminus of the KS domain and the *C*-terminus of the adapter domain, respectively.
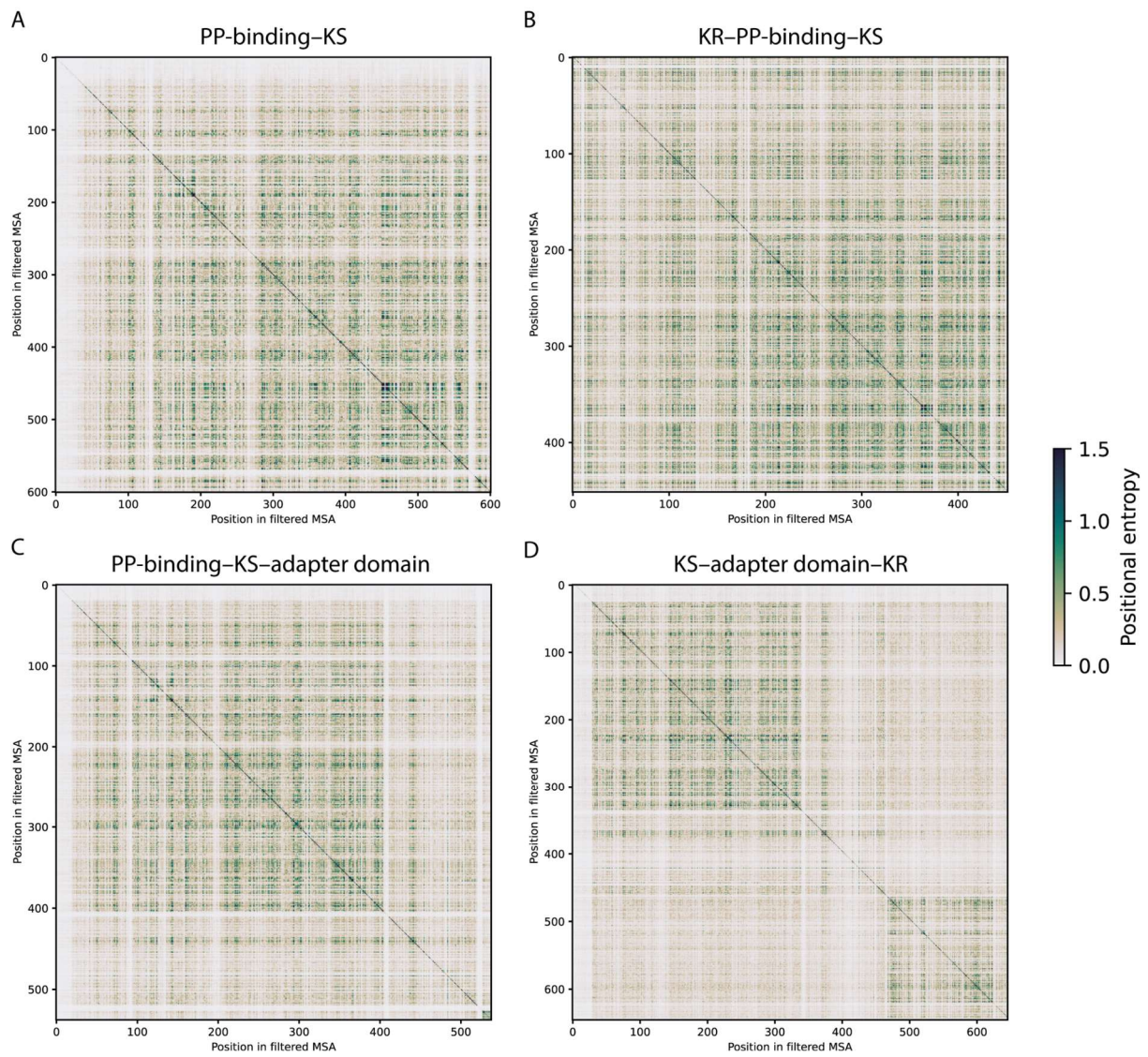
Figure S10. SCA matrices of conservation-filtered alignments obtained using the MUSCLE algorithm of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR.
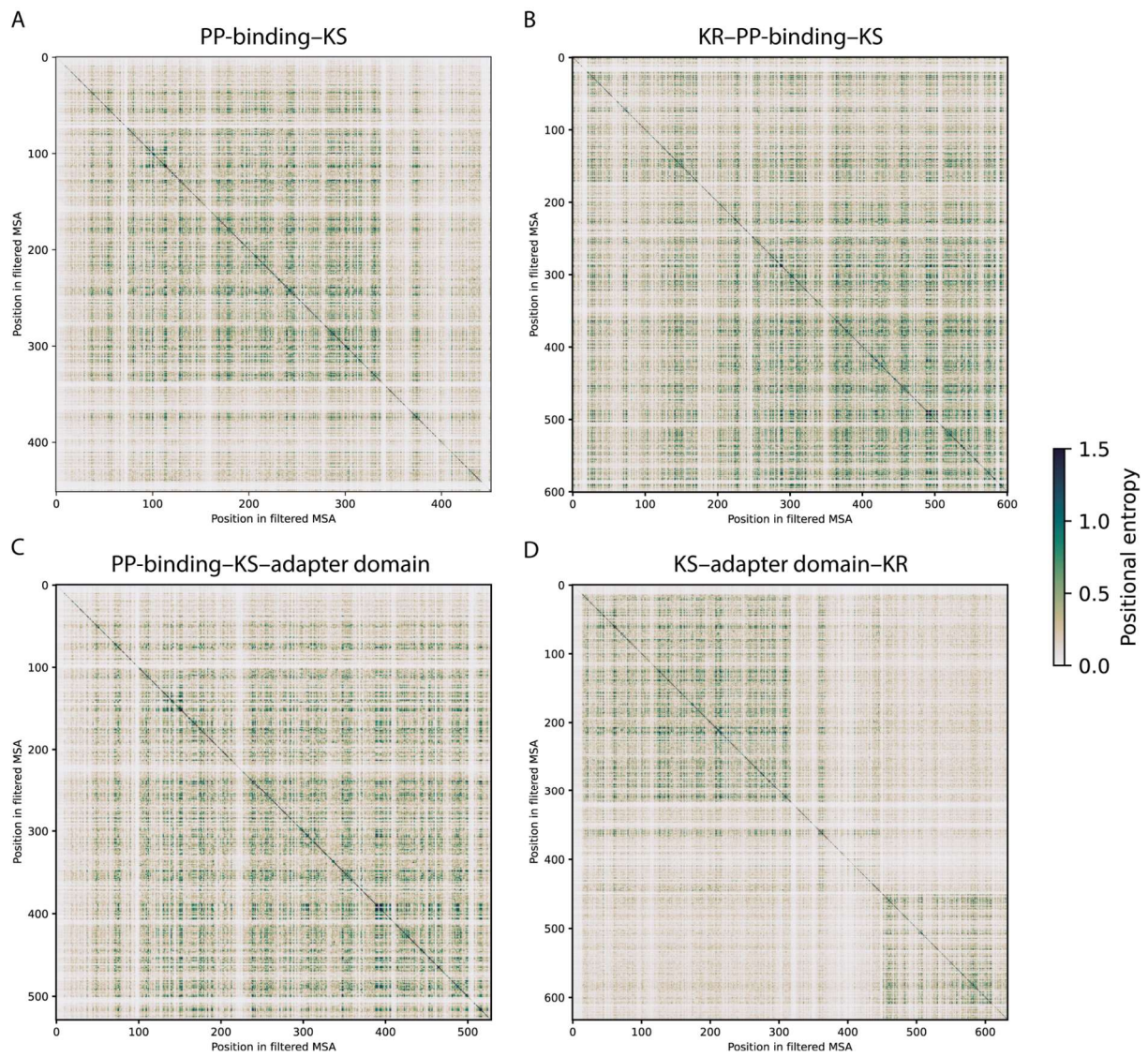
Figure S11. SCA matrices of conservation-filtered alignments obtained using the Clustal-Omega algorithm of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR.
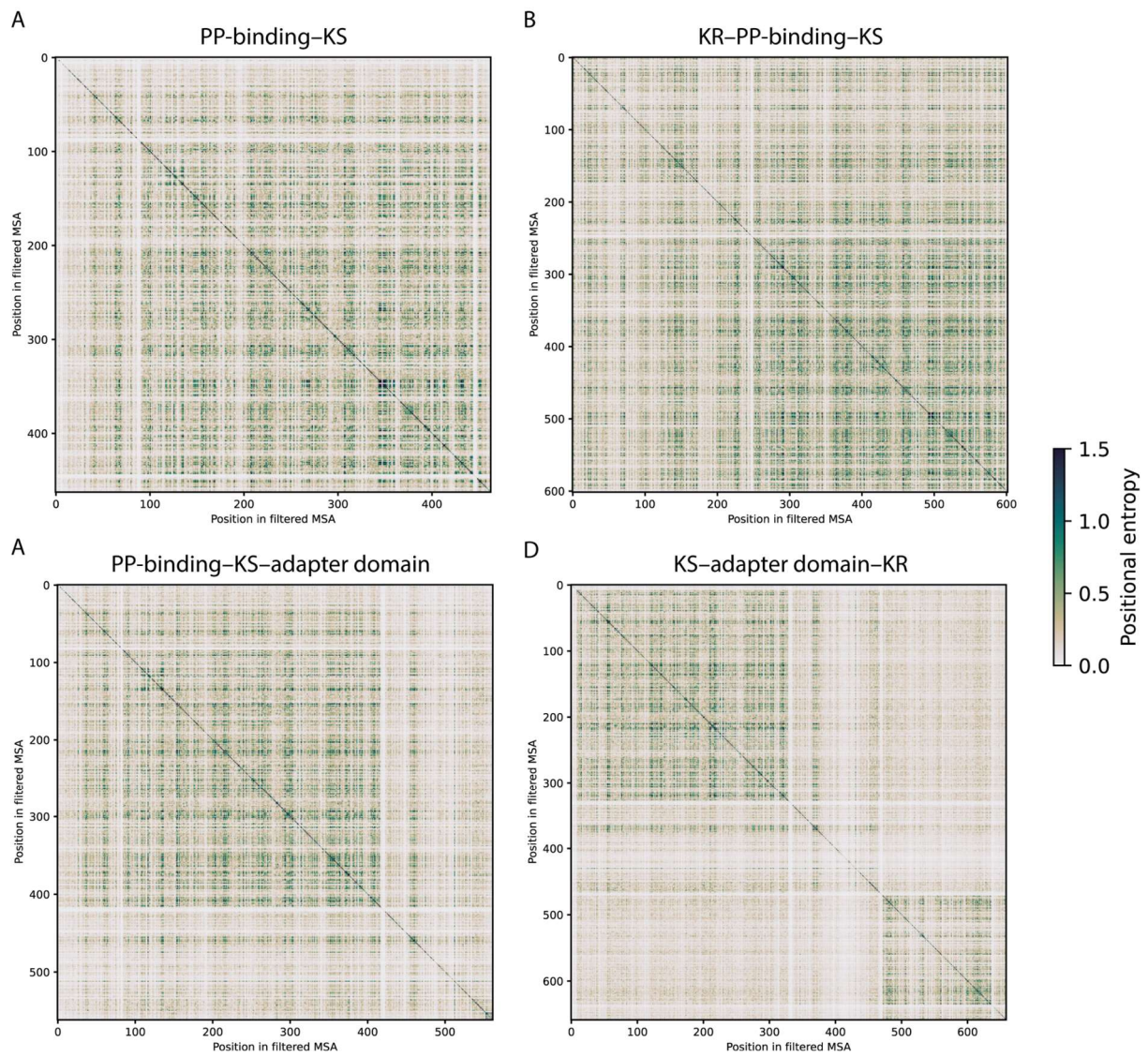
Figure S12. SCA matrices of conservation-filtered alignments obtained using the MAFFT algorithm and an ep offset of 0.0 of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR.
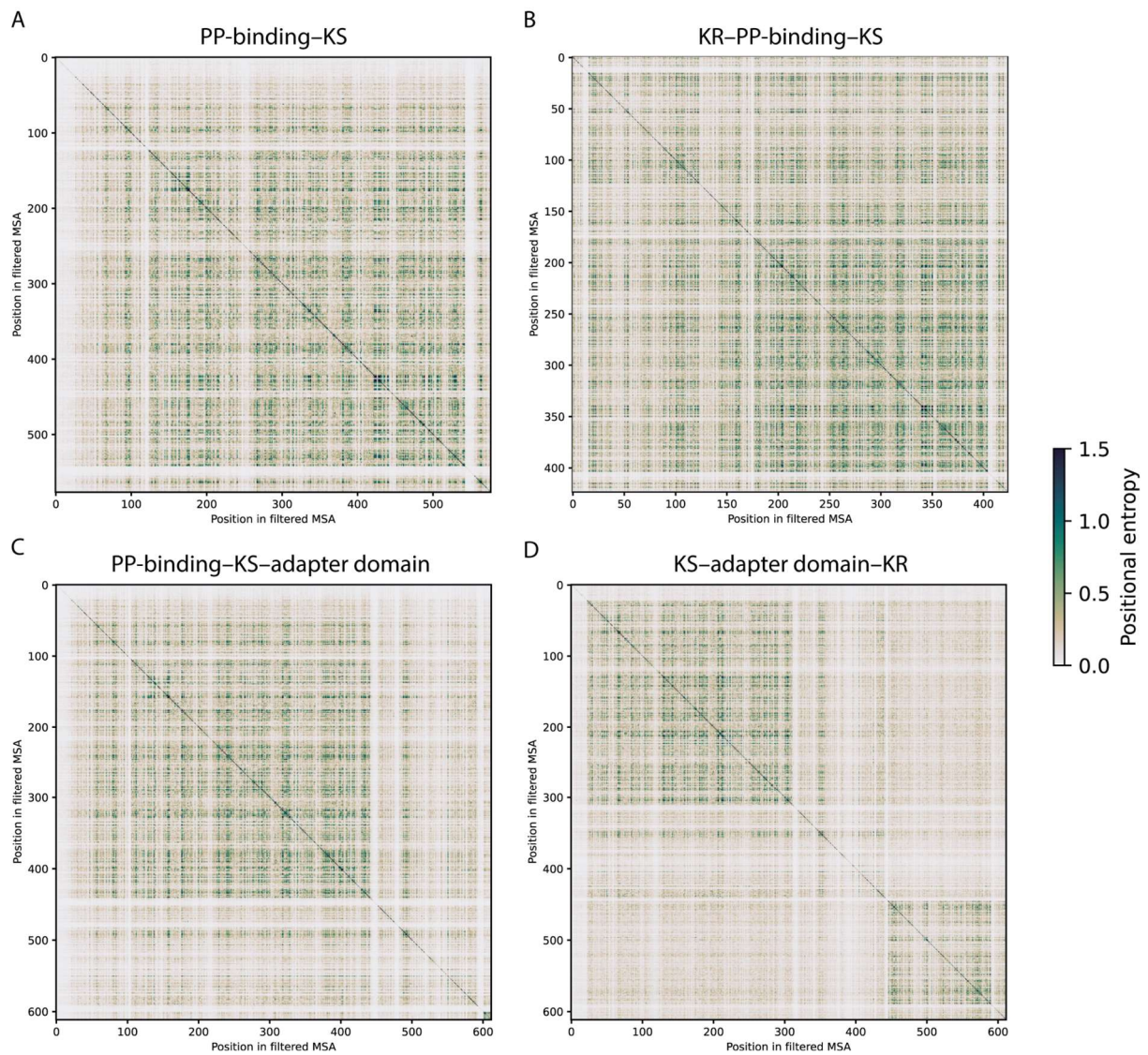
Figure S13. SCA matrices of conservation-filtered alignments obtained using the MAFFT algorithm and an ep offset of 0.123 of A) PP-binding–KS, B) KR–PP-binding–KS, C) PP-binding–KS–adapter domain and D) KS–adapter domain–KR.

188

MUSCLE — KS–adapter domain–KR

PKPAEPRRKRPAKEPAPAAAKREAR-EPIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWR
AYYPDKKGKTYCKWGGFLDGIDEFDPLFFNISPREAELMDPQQRLFLQEAWKALEDAGYTPESLSGKK
TGVFVGVMNNGYGLLGRAHAAGNSPSIANRISYFLNLKGPSIPVDTACSSSLVAIHLACQALRNGECDM
ALAGGVNLYLTPEKYISLSKAGMLSPDGRCKTFDAGADGFVPGEGVGAVLLKPLSDAEADGDHIYGVI
RGSAINHGGKTNGITAPNPKAQADLIRDAYRKAGIDPETISYIEAHGTGTKLGDPIEINGLSRAFREKTDR
KQFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHFETLNPHIDFDDSPFYVNTELQPWEDGD
GPRRAGVSSFGAGGTNAHIVIEEYPKAR-PEEPALIVLSAKNEERLKEYAEQLLDF-IDLADLAYTLQTGR
EAMEERLAFVAGSREELEEKLEAFLDG-GGYRGKETLATADEDMAALVDAWIEKGKLAKLAELWVKG
LDIDWNRLYGGEKPRRISLPTYPFAKERYWPE-PFPEDGVYLITGGTGGLGLLCARHLAERVKKLVLTG
RELAEKIRALRELEKGAQVEYLSADLTDEAAVRQALEQIKRTFGPIGGVIHCAGVVEDPAFIRKTAEDIQ
RVLEPKVAGLQTLDEALANEPLDFFVLFSSVSAVIPAAGQSDYAMANAFMDYFAAARGPIVSIQWPNW
KETGMEVTSKALKQSGLLSLTNAEGLALLDR

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14   Sector 15   Sector 16

MUSCLE — PP-binding–KS

---PSEPA-AAPAPKAAE-ASEALLEKALEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKK
FGELSPTLFFEYPTIAELAEYLLEEYPAADEIAIIGMSGRFPGA-DLEEFWENLKEGKDCITEIPAERWWRE
GKTYSKWGGFIDGIDEFDPLFFGISPREAELMDPQQRLFLEEAWKAIEDAGYTPEELSGKKVGVFVGVM
SGDYQLLAEAKNIVATGQSSIANRVSYFFNLHGPSEAVDTACSSSLVAIHLACQSLRNGECEMAIAGGV
NLLLHPEKYIMLSQAGMLSPDGRCKTFDERADGYVPGEGVGAVLLKPLSKAEADGDHIYAVIKGSAVN
HGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYIEAHGTGTSLGDPIEINALTKAFDDPCGIGSVKSN
IGHLESAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKSPFYVQRLEEWGGNIRRAGISSFGAGGTN
AHVILEYE--Q-RPLIVLSAKNEERLKEYA

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
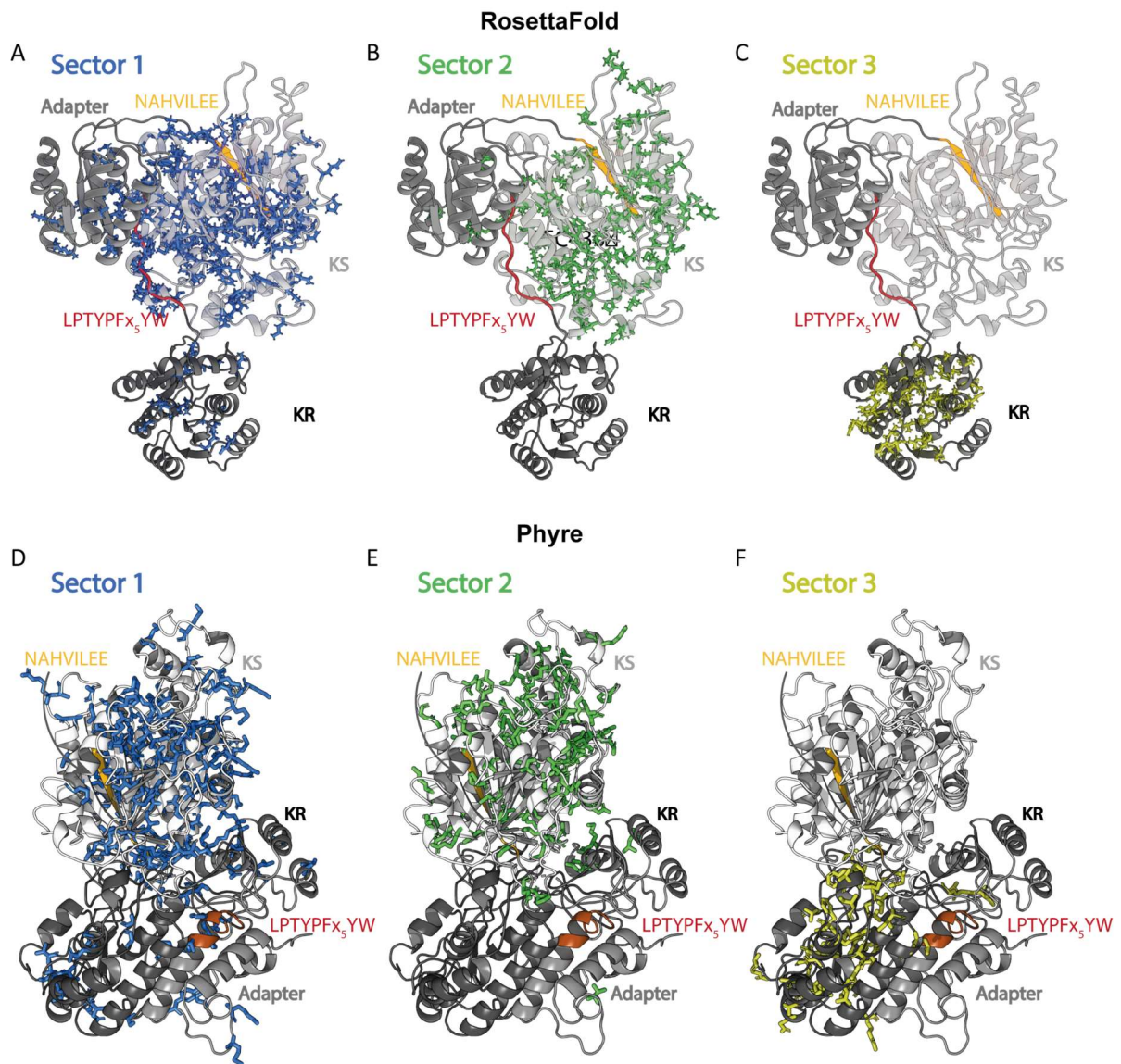Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12

MUSCLE — PP-binding–KS–adapter domain

EYSTEEAAAELAEPAAESEGEKLEDYLKKLVSEVLKIPAEDIDADTPLEEYGFDSISLTELTNRLNEEFGELSPTLFFE
YQTIAELAAYLLEE-AADEPIAIIGISGRFPGADDLEEFWENLKEGKDCITEIPEDRWDWRAYGDPDEPGKTYSKWGG
FIDGVDEFDPLFFGISPREAELMDPQERLFLETAWKAIEDAGYTPESLAGKVGVFVGVMNGDYQLLGAE--GYAATG
LSPSSIANRVSYFLNLHGPSEAVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSKAGMLSSDGRC
KTFGKGADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGIDP
RTISYIEAHGTGTSLGDPIEINGLKKAFRTDFCAIGSVKSNIGHLESAAGIAGLIKVLLQMKHKTLVPSLHSEELNPYID
FEDSPFYVQQELQEWKRPELPRRAGVSSFGAGGVNAHVVLEEYIPEAADGPALIVLSAKNEERLREYARRLLDFLLA
DIAYTLQVGREAMEERLAFVASSLEELEEKLEAFLEGVFRGQVKRNKF-DEDMEIAWLEKGKLAKLAELWVKGADI
DWNRLYKPRRISLPTYPFARERYWVP----ASLHPLLHENTST

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13

Figure S14. Color-coded consensus sequences of various domain motifs. To visualize the protein sectors obtained from the SCA on the consensus sequence of the MSA, color-coded consensus sequences are given below, with the color of each residue corresponding to the sector it is in. Residues that are not in a protein sector are colored black.

MUSCLE — KR–PP-binding–KS

VCYRNGRYVAYLK-P-PFKDDGVYLITGGTGGLGLLFARHLAERGVKKLVLTGRSLREADKIEAIQELE
AKGVQVEYLSADLTDREAVERLIEEIKRTFGPIGGVIHCAGVADDPAFIRKTLEDIQEVLEPKVAGLQNL
DEATKDEPLDFFVLFSSVSAAIGNAGQSDYAMANAFMDYFAAYRNAGGKTLSINWPLWKEGGMGEK
VEEWLKELFSEELKIPAEQLDTDEPFQDYGVDSILLAQLAQRLNKKFGDLDPTVFFEYPTIESLAEYLLSE
YASTEDIAIIGMSCRFPGADDLEEFWDNLREGRDAITEVPAERWGKSYSKWGGFIDGIDQFDPLFFNISPR
EAELMDPQERLFLEESWKALEDAGYTREELSGKKVGVFVGGRSGHYPLLGLEQGKNPIVGGGQNYIAA
RVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECEAAVAGGVNLLLDPSKYLMLSQAGMLSPDGR
CHTFDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPTAPNPEAQAEVIKSAL
EKAGIDPETISYIEAHGTGTSLGDPIELKALTKVFR-RTDKQFCGIGSVKSNIGHLESAAGIASLIKVLLQL
KHKQLPPSLHCEEPNPYIDFEKSPFYVNRELEEWREPRRAGISSFGFGGTNAHVIVEEYEEDERKPLLIVL
SAKTEERLKEKA-

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14   Sector 15   Sector 16   Sector 17   Sector 18
Sector 19   Sector 20   Sector 21   Sector 22

Clustal-Omega — KS–adapter domain–KR

-P-AAAAEPADRRAEPIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPKGKTYCKWGGFLDGI
DEFDPLFFNISPREAELMDPQQRLFLQEAWKALEDAGYTPESLSKRTGVFVGVMNNGYGLLLEGHAATGNFSIIANR
ISYFLNLKGPSIPVDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGMLSPDGRCKTFDAGAD
GYVPGEGVGAVLLKPLSDAEADGDHIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYEKAGIDPETISYIEAHGT
GTKLGDPIEINGLSRAFRELQFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHFETLNPHIDFDDSPFYVN
TELQPWERP--PRRAGVSSFGAGGTNAHIVIEEYQPKAR-PEEPALIVLSAKNEERLKEYAEQLLDFLRA--DIDLADLA
YTLQTGREAMEERLAFVAGSREELEEKLEAFLDGK-GIYRGQVKREAADEDMAALVDAWIRKGKLAKLAELWVK
GLDIDWNRLYGGKPRRISLPTYPFAKERYWLPEPPFPEDGVYLITGGTGGLGLLCARHLAERYGVKKLVLTGREPLP
PLAEIQELEAKGVQVEYLSADLTDEAAVRQALEQIKMGPIGGVIHCAGVTD-NAFIRKTAEDIQRVLEPKVAGLQTL
DEALANPLDFFVLFSSVSAVIPAAGQSDYAMANAFMDYFAAARQ-GPIVSIQWPNWKETGMGEV-AYKQSGLLSLT
NAEGLQLLDR

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14

Clustal-Omega — PP-binding–KS

EEEASAADLLEKLEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKLLSPTLFFEYPTIAELAEYLLE
EYP-ADEDIAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDWRETYSKWGGFIDGIDEFDPLFFGISPREAEL
MDPQQRLFLEEAWKAIEDAGYTPEELSGKVGVFVGVMSGDYQLLLAEAAYNATGLSQSSIANRVSYFFNLHGPSEA
VDTACSSSLVAIHLACQSLRNGECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKTFDERADGYVPGEGVGAVL
LKPLSKAEADGDHIYAVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYIEAHGTGTSLGDPIEINA
LTKAFRELKQFCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLCEELNPYIDFEKSPFYVQRELEEWKRP-PR
RAGISSFGAGGTNAHVILEEYEPP---P-LIVLSAKNEERLREY

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13

Figure S14 continued.

Clustal-Omega — PP-binding–KS–adapter domain

-----ESDAEDLREKLEDYLKKLVSEVLKIPAEDIDADTPLEEYGFDSISLTELTNRLNEEFGLSPTLFFEYQTIAELAAYL
LEEH-AADEPIAIIGISGRFPGADDLEEFWENLKEGKDCITEIPEDRWDWREYYGDGKTYSKWGGFIDGVDEFDPLFF
GISPREAELMDPQERLFLETAWKAIEDAGYTPESLAKVGVFVGVMGDYQLLGAEEEHAATGSSPSSIANRVSYFLNL
HGPSEAVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSKAGMLSSDGRCKTFGKGADGYVPGEG
VGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGIDPRTISYIEAHGTGTSLGD
PIEINGLKKAFRQLKQFCAIGSVKSNIGHLESAAGIAGLIKVLLQMKHKTLVPSLHSEELNPYIDFEDSPFYVQQELQE
WKRPVELPRRAGVSSFGAGGVNAHVVLEEYIPPAAGGPALIVLSAKNEERLREYARRLLDFLER-DADLADIAYTLQ
VGREAMEERLAFVASSLEELEEKLEAFLDGEK-VFRGQVKRNMDEAIDAWLEKGKLAKLAELWVKGAEIDWNRLY
GPRRISLPTYPFARERYWVPEKESKAA--HE

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14   Sector 15

Clustal-Omega — KR–PP-binding–KS

VCYRNGKRYVAYKEIP-SGSVPFKDDGVYLITGGTGGLGLLFARHLAERYGAKLVLTGRSPLPPKEKI-ELEAKGVQ
VEYLSADLTDREAVERLIEEIKRTFGPIGGVIHCAGVADDSFIRKTLEDIQEVLEPKVAGLQNLDEATKDEPLDFFVLF
SSVSAVIGNAGQSDYAMANAFMDYFAAYNRQGKTLSINWPLWKEGGMGEEVEEWLKELFSEELKIPAEQLDTDEP
FQDYGVDSILLAQLAQRLNKKFG-LDPTVFFEYPTIESLAEYLLTEYPSAEDIAIIGMSCRFPGADDLEEFWDNLREGR
DAITEVPAERWDWRSSKWGGFIDGIDQFDPLFFNISPREAELMDPQERLFLEESWKALEDAGYTREELSGKKVGVFV
GGRSGHYPLLGALESKNPIVGGGQNYIAARVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECEAAVAGGVNL
LLHPSKYLMLSQAGMLSPDGRCHTFDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPT
APNPEAQAEVIKSALEKAGIDPETISYIEAHGTGTSLGDPIEIKALTKVFRSDDKQFCGIGSVKSNIGHLESAAGIASLIK
VLLQLKHKQLPPSLHCEEPNPYIDFEKSPFYVNRELEEWKRAPRRAGISSFGFGGTNAHVIVEEYEESAKLSAKTLKE
YAI

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14   Sector 15   Sector 16   Sector 17

MAFFT with ep = 0.0 — KS–adapter domain–KR

PKEPEPEAIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPGKTYCKWGGFLDGIDEFDPLFFN
ISPREAELMDPQQRLFLQEAWKALEDAGYTPESLGRTGVFVGVMNNGYQLLGAATGNSFSIIANRISYFLNLKGPSIP
VDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGMLSPDGRCKTFDAGADGYVPGEGVGAV
LLKPLSDAEADGDHIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYEKAGIDPETISYIEAHGTGTKLGDPIEING
LSRAFRELTDDKFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHFETLNPHIDFDDSPFYVNTELQPWDG
-PRRAGVSSFGAGGTNAHVIVEEYPAR-PE-PALIVLSAKNEERLKEYAEQLLDFLRIDLADLAYTLQTGREAMEERLA
FVAGSREELEEKLEAFLDG-NEGIYRGQVKRNKDTLALFTADEDMAALVDAWIEKGAKLAELWVKGLDIDWNRLY
PRRISLPTYPFAKERYWLP-PED-FKEGGVYLITGGTGGLGLLCARHLAERYGVKKLVLTGREPLPKVKAIQALEAKG
VQVEYLSADLTDEAAVRQALEQIKQTMGPIGGVIHCAGVVENPAFIRKTAEDIQRVLEPKVAGLQTLDEALANEPLD
FFVLFSSVSAVIPAGQSDYAMANAFMDYFAIVSIQWPNWKETGMVRTKALKSGLLSLTNAEGLALLD

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14   Sector 15

Figure S14 continued.

MAFFT with ep = 0.0 — PP-binding–KS

ASSPAA-EKVEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKFGELSPTLFFEYPTIAELAEYLLEE
YPA-AEDIAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDWREYSKGGFIDGIDEFDPLFFGISPREAELMDP
QQRLFLEEAWKAIEDAGYTPEELSKKVGVFVGVMNGDYQLAEA-ATG-SQSIANRVSYFFNLHGPSEAVDTACSSSL
VAIHLACQSLRNGECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKTFDERADGYVPGEGVGAVLLKPLSKAE
ADGDHIYGVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYIEAHGTGTSLGDPIEINALTKAFRS-D
KCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKSPFYVQRELEEWKGPRRAGISSFGAGG
TNAHVILEEYEAR-PLSAKNERLREYA

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11

MAFFT with ep = 0.0 — PP-binding–KS–adapter domain

-E-LREKVELKKLVSEILKIPAEDIDADTPLEEYGFDSISLTELTNRLNEEFGELSPTLFFEYQTIAELAAYLLEEHPAEDI
AIIGISGRFPGADDLEEFWENLKEGKDCITEIPEERWDWREYYDPPGKTYSKWGGFIDGVDEFDPLFFGISPREAELM
DPQERLFLETAWKAIEDAGYTPESLAKVGVFVGVMNGDYQLLGE-ATGLSPSSIANRVSYFLNLHGPSEAVDTACSS
SLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSQAGMLSSDGRCKTFGKGADGYVPGEGVGAVLLKPLSKA
EADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGIDPRTISYIEAHGTGTSLGDPIEINGLKKAFRQ
LTDKFCAIGSVKSNIGHLESAAGIAGLTKVLLQMKHKTLVPSLHSEELNPYIDFEDSPFYVQQELQEWKRG-PRRAGV
SSFGAGGVNAHVVLEEYPPA--GPALIVLSAKNEERLREYARRLLDFLEQ-DLADIAYTLQVGREAMEERLAFVASSL
EELEEKLEAFLDGEKEGVFRGQVKRNADEDMDETIEAWLEKGKLAKLAELWVKGAEIDWNRLYKPRRISLPTYPF
ARERYWVPHPLLHENT

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13

MAFFT with ep = 0.0 — KR–PP-binding–KS

VCYRNGKRYVAYLKE-PFKDGVYLITGGTGGLGLLFARHLAERYVAKLVLTGRSPLTSKKEKIKELEAKGVQVEYL
SADLTDREAVERLIEEIKRTFGPIGGVIHCAGVADDPAFIRKTLEDIQEVLEPKVAGLQNLDEATKDEPLDFFVLFSSV
SAVIGNLGQSDYAMANAFMDYFAAYRKGQKTLSINWPLWKEGGMGEKVELKELFSEELKIPAEQLDTDEPFQDYG
VDSILLAQLAQRLNKKFGDLDPTVFFEYPTIESLAEYLLSEYPASEEIAIIGMSCRFPGADDLEEFWDNLREGRDAITE
VPAERWDWSNSKWGGFIDGIDQFDPLFFNISPREAELMDPQERLFLEESWKALEDAGYTREELSGKKVGVFVGGRS
GHYPLEERGKPIVGNGQNIAARVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECEAAVAGGVNLLLHPSKYL
MLSQAGMLSPDGRCHTFDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPTAPNPEAQA
EVIKSALEKAGIDPETISYIEAHGTGTSLGDPIEIKALTKVFDKQFCGIGSVKSNIGHLESAAGIASLIKVLLQLKHKQL
PPSLHCEEPNPYIDFEKSPFYVNRELEEWEPRRAGISSFGFGGTNAHVIVEEYRKPLVLSKERLESAE

Sector 1 Sector 2 Sector 3 Sector 4 Sector 5 Sector 6
Sector 7 Sector 8 Sector 9 Sector 10   Sector 11   Sector 12
Sector 13   Sector 14   Sector 15   Sector 16   Sector 17

Figure S14 continued.

192

MAFFT with ep = 0.123 — KS–adapter domain–KR

APKPASRKPRPAAKPEPKAAAADEEPIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPKKGK
TYCKWGGFLDGIDEFDPLFFNISPREAELMDPQQRLFLQEAWKALEDAGYTPESLSGRTGVFVGVMNNGYGLLGNE
QAEHAATGNSPSIANRISYFLNLKGPSIPVDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGM
LSPDGRCKTFDAGADGYVPGEGVGAVLLKPLSDAEADGDHIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYE
KAGIDPETISYIEAHGTGTKLGDPIEINGLSRAFREYTDRKQFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVP
SLHFETLNPHIDFDDSPFYVNTELQPWERDGAPRRAGVSSFGAGGTNAHIVIEEYQPKARAPEEPALIVLSAKNEERL
KEYAEQLLDFLRANIDLADLAYTLQTGREAMEERLAFVAGSREELEEKLEAFLDGKTNEGIYRGQVKRNKDTIALF
EADEDMAALLDAWIRKGKLAKLAELWVKGLDIDWNRLY-GEKPRRISLPTYPFAKERYWLP-PEDPFKEGGVYLIT
GGTGGLGLLCARHFAERYGVKKLVLTGREPLPPRKEAAIRELEAKGVQVEYLSADLTDEAAVRQALEQIKRTMGPI
GGVIHCAGVEDPAFIRKTAEDIQRVLEPKVAGLQTLDEALANEPLDFFVLFSSVSAVIGSAGQSDYAMANAFMDYF
AAIVSIQWPNWKETGMEVETKAYQEGLLSLTNAEGLQLLDR

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
        Sector 7   Sector 8   Sector 9   Sector 10 Sector 11   Sector 12
        Sector 13   Sector 14

MAFFT with ep = 0.123 — PP-binding–KS

--K-AAAPAQEEAPAAESDADL-EKVEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKFGELSPTLF
FEYPTIAELAEYLLEEYPAAAEDIAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDGGKTYSKWGGFIDGID
EFDPLFFGISPREAELMDPQQRLFLEEAWKAIEDAGYTPEELSGKKVGVFVGVMHGDYQLLL-AGNPIAATGN--SIA
NRVSYFFNLHGPSEAVDTACSSSLVAIHLACQSLRNGECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKTFDER
ADGYVPGEGVGAVLLKPLSKAEADGDHIYAVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYIEA
HGTGTSLGDPIEINALTKAFR-LTDDKQCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKS
PFYVQRELEEWPDGPRRAGISSFGAGGTNAHVILEEY--A-A--RRKP-LIVLSAKNEERKEYA

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
        Sector 7   Sector 8   Sector 9   Sector 10 Sector 11   Sector 12

MAFFT with ep = 0.123 — PP-binding–KS–adapter domain

--AES-ADA--AASEEEEASLREKVEAYLKKLVSEVLKIPAEEIDADTPLEEYGFDSISLTELTNRLNEEFGELSPTLFFEY
QTIAELAAYLLEEH-AADDIAIIGISGRFPGADDLEEFWENLKEGKDCITEIPEDRWDWRAYYDPD-PGKTYSKWGGF
IDGVDEFDPLFFGISPREAELMDPQERLFLETAWKAIEDAGYTPESLATKVGVFVGVMNGDYQLLGEEG-AATGNYS
SIANRVSYFLNLHGPSEAVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSQAGMLSSDGRCKTFG
KGADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGIDPRTISY
IEAHGTGTSLGDPIEINGLKKAFRQLTDKQFCAIGSVKSNIGHLESAAGIAGLIKVLLQMKHKTLVPSLHSEELNPYID
FEDSPFYVQQELQEWKRNPRRAGVSSFGAGGVNAHVVLEEYIPPAAAA-DGPALIVLSAKNEERLREYARRLLDFLE
RE--TDLADIAYTLQVGREAMEERLAFVASSLEELEEKLEAFLDGEK-EGVFRGQVKRNKEALA--DE-DELIEAWLEK
GKLAKLAELWVKGAEIDWNRLYGRKPRRISLPTYPFARERYWVP------RLHPLLHENTS

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
        Sector 7   Sector 8   Sector 9   Sector 10   Sector 11   Sector 12
        Sector 13   Sector 14

Figure S14 continued.

193

MAFFT with ep = 0.123 — KR–PP-binding–KS

VCYRNGKRYVAYLKEIP---S-VPFKDDGVYLITGGTGGLGLLFARHLAERGVAKLVLTGRSPLTSKKEKIKELEAKG
VQVEYLSADLTDREAVERLIEEIKRTFGPIGGVIHCAGVADDPAFIRKTLEDIQEVLEPKVAGLQNLDEATKDEPLDF
FVLFSSVSAVIGNAGQSDYAMANAFMDYFAAYRKGQGKTLSINWPLWKEGGMGEKVEEWLKELFSEELKIPAEQL
DTDEPFQDYGVDSILLAQLAQRLNKKFGDLDPTVFFEYPTIESLAEYLLSEYPASEEIAIIGMSCRFPGADDLEEFWDN
LREGRDAITEVPAERWDWSNSKWGGFIDGIDQFDPLFFNISPREAELMDPQARLFLEESWKALEDAGYTREELSGKK
VGVFVGGRSGHYPDEERGNPIVGNGQNIAARVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECEAAVAGGVN
LLLHPSKYLMLSQAGMLSPDGRCHTFDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPT
APNPEAQAEVIKSALEKAGIDPETISYIEAHGTGTSLGDPIELKALTKVFTDDKQFCGIGSVKSNIGHLESAAGIASLIK
VLLQLKHKQLPPSLHCEEPNPYIDFEKSPFYVNRELEEWEPRRAGISSFGFGGTNAHVIVEEYEDEQ--KRTPLPPELA
KTEERLDEYAE

Sector 1   Sector 2   Sector 3 Sector 4   Sector 5   Sector 6   Sector 7
        Sector 8   Sector 9   Sector 10   Sector 11   Sector 12   Sector 13
        Sector 14   Sector 15

Color coded consensus sequences of various domain motifs from conservation-filtered alignments

The protein sectors were also mapped on the consensus sequence of the conservation-filtered MSAs. In addition to the color coding, the highly conserved (>80%) positions that are filtered out for the MSA are italicized in the consensus sequence. These positions are not used in the SCA and hence do not show up in protein sectors.

MUSCLE — KS–adapter domain–KR — Conservation-filtered

PEAPPPKPVRKKRFAEPEPAK-AD--
APIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPKKGKTYCKWGGFLDGIDEFDPLFFNISPR
EAELMDPQQRLFLQEAWKALEDAGYTPESLSGKNTGVFVGVMNNGYGLLLNRGGGHAATGNSPSIANRISYFLNLK
GPSIPVDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGMLSPDGRCKTFDAGADGYVPGEGV
GAVLLKPLSDAEADGDTIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYEKAGIDPETISYIEAHGTGTKLGDPIEI
NGLSRAFREYTDDKQFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHFETLNPHIDFDDSPFYVNTELQP
WERDGPRRAGVSSFGAGGTNAHIVIEEYQPKA-
APPALIVLSAKNEERLKEYAEQLLDIDLADLAYTLQTGREAMEERLAFVAGSREELEEKLEAF-
GEEYRGQVKRNKDTIAFTADEDMAELVDAWIEKGKYAKLAELWVKGLDIDWNRLYGGEKPRRISLPTYPFAKERY
WVPEPEFPEDGVYLITGGTGGLGLLCARHLAERYGVKKLVLTGREPLPRSERELEAKGVQVEYLSADLTDEAAVRQ
ALEQIKRTMGPIGGVIHCAGVLENPAFIRKTAEDIQRVLEPKVAGLQTLDEALANEPLDFFVLFSSVSAVIPAAGQSDY
AMANAFMDYFAAARQGPIVSIQWPNWKETGMEVTSKALKQSGLLSLTNAEGLALLDR

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
        Sector 7   Sector 8   Sector 9   Sector 10   Sector 11   Sector 12

MUSCLE — PP-binding–KS — Conservation-filtered

---PSEPA-AAPAPKAAE-
ASEALLEKALEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKFGEISPTLFFEYPTIAELAEYLLEEY
PAADELAIIGMSGRFPGA-
DLEEFWENLKEGKDCITEIPAERWWREGKTYSKWGGFIDGIDEFDPLFFGISPREAELMDPQQRLFLEEAWKAIEDA
GYTPEELSGKKVGVFVGVMSGDYQLLAEAKNIVATGQSSIANRVSYFFNLHGPSEAVDTACSSSLVAIHLACQSLRN
GECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKTFDERADGYVPGEGVGAVLLKPLSKAEADGDHIYAVIKGS
AVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYIEAHGTGTSLGDPIEINALTKAFDDPCGIGSVKSNIGHLE
SAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKSPFYVQRLEEWGGNIRRAGISSFGAGGTNAHVILEYE--Q-
RPLIVLSAKNEERLKEYA

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
        Sector 7   Sector 8   Sector 9   Sector 10

Figure S14 continued.

MUSCLE — PP-binding–KS–adapter domain — Conservation-filtered

EYSTEEAAAELAEPAAESEGEKLEDYLKKLVSEVLKIPAEDIDADTPLEEYGFDSISLTELTNRLNEEFGELSPTLFFE
YQTIAELAAYLLEE-
AADEPIAIIGISGRFPGADDLEEFWENLKEGKDCITEIPEDRWDWRAYGDPDEPGKTYSKWGGFIDGVDEFDPLFFGI
SPREATLMDPQERLFLETAWKAIEDAGYTPESLAGKVGVFVGVMNGDYQLLGAE--
GYAATGLSPSSIANRVSYFLNLHGPSEAVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSKAGMLSS
DGRCKTFGKGADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKA
GIDPRTISYIEAHGTGTSLGDPIEINGLKKAFRTDFCAIGSVKSNIGHLESAAGIAGLIKVLLQMKHKTLVPSLHSEELNP
YIDFEDSPFYVQQELQEWKRPELPRRAGVSSFGAGGVNAHVVLEEYIPEAADGPALIVLSAKNEERLREYARRLLDFL
LADIAYTLQVGREAMEERLAFVASSLEELEEKLEAFLEGVFRGQVKRNKF-
DEDMEIAWLEKGKLAKLAELWVKGADIDWNRLYKPRRISLPTYPFARERYWVP----ASLHPLLHENTST

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
    Sector 7   Sector 8   Sector 9   Sector 10   Sector 11

MUSCLE — KR–PP-binding–KS — Conservation-filtered

VCYRNGRYVAYLK-P-
PFKDDGVYLITGGTGGLGLLFARHLAERGVKKLVLTGRSLREADKIEAIQELEAKGVQVEYLSADLTDREAVERLIE
EIKRTFGPIGGVIHCAGVADDPAFIRKTLEDIQEVLEPKVAGLQNLDEATKDEPLDFFVLFSSVSAAIGNAGQSDYAMA
NAFMDYFAAYRNAGGKTLSINWPLWKEGGMGEKVEEWLKELFSEELKIPAEQLDTDEPFQDYGVDSILLAQLAQRL
NKKFGDLDPTVFFEYPTIESLAEYLLSEYASTEDIAIIGMSCRFPGADDLEEFWDNLREGRDAITEVPAERWGKSYSK
WGGFIDGIDQFDPLFFNISPREAELMDPQERLFLEESWKALEDAGYTREELSGKKVGVFVGGRSGHYPLLGLEQGKN
PIVGGGQNYIAARVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECFAAVAGGVNLLLDPSKYLMLSQAGMLSP
DGRCHTFDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPTAPNPEAQAEVIKSALEKAGI
DPETISYIEAHGTGTSLGDPIELKALTKVFR-
RTDKQFCGIGSVKSNIGHLESAAGIASLIKVLLQLKHKQLPPSLHCEEPNPYIDFEKSPFYVNRELEBWREPRRAGISSF
GFGGTNAHVIVEEYEEDERKPLLIVLSAKTEEBLKEKA-

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6
    Sector 7   Sector 8   Sector 9   Sector 10   Sector 11

Clustal-Omega — KS–adapter domain–KR — Conservation-filtered
-P-
AAAAEPADRRAEPIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPKGKTYCKWGGFLDGIDE
FDPLFFNISPREAELMDPQQRLFLQEAWKALEDAGYTPESLSKRTGVFVGVMNNGYGLLLEGHAATGNFSIIANRISY
FLNLKGPSIPVDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGMLSPDGRCKTFDAGADGYVP
GEGVGAVLLKPLSDAEADGDHIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYEKAGIDPETISYIEAHGTGTKLG
DPIEINGLSRAFRELQFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHFETLNPHIDFDDSPFYVNTELQPW
ERP--PRRAGVSSFGAGGTNAHVIEEYQPKAR-PEEPALIVLSAKNEERLKEYAEQLLDFLRA--
DIDLADLAYTLQTGREAMEERLAFVAGSREELEEKLEAFLDGK-
GIYRGQVKREAADEDMAALVDAWIRKGKLAKLAELWVKGLDIDWNRLYGGKPRRISLPTYPFAKERYWLPEPPFPE
DGVYLITGGTGGLGLLCARHLAERYGVKKLVLTGREPLPPLAEIQELEAKGVQVEYLSADLTDEAAVRQALEQIKM
GPIGGVIHCAGVTD-
NAFIRKTAEDIQRVLEPKVAGLQTLDEALANPLDFFVLFSSVSAVIPAAGQSDYAMANAFMDYFAAARQ-
GPIVSIQWPNWKETGMGEV-AYKQSGLLSLTNAEGLQLLDR

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6   Sector 7
    Sector 8   Sector 9   Sector 10   Sector 11   Sector 12

Figure S14 continued.

Clustal-Omega — PP-binding–KS — Conservation-filtered

EEEASAADLLEKLEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKLLSPTLFFEYPTIAELAEYLLE
EYP-
ADEDIAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDWRETYSKWGGFIDGIDEFDPLFFGISPREAELMDP
QQRLFLEEAWKAIEDAGYTPEELSGKVGVFVGVMSGDYQLLLAEAAYNATGLSQSSIANRVSYFFNLHGPSEAVDT
ACSSSLVAIHLACQSLRNGECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKTFDERADGYVPGEGVGAVILKPL
SKAEADGDHIYAVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYIEAHGTGTSLGDPIEINALTKAF
RELKQFCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLCEELNPYIDFEKSPFYVQRELEEWKRP-
PRRAGISSFGAGGTNAHVILEEYEPP---P-LIVLSAKNEERLREY

Sector 1　Sector 2　Sector 3 Sector 4　Sector 5　Sector 6　Sector 7
　　Sector 8　　Sector 9　　Sector 10

Clustal-Omega — PP-binding–KS–adapter domain — Conservation-filtered

-----
ESDAEDLREKLEDYLKKLVSEVLKIPAEDIDADTPLEEYGFDSISLTELTNRLNEEFGLSPTLFFEYQTIAELAAYLLEE
H-
AADEPIAIIGISGRFPGADDLEEFWENLKEGKDCITEIPEDRWDWREYYGDGKTYSKWGGFIDGVDEFDPLFFGISPR
EAELMDPQERLFLETAWKAIEDAGYTPESLAKVGVFVGVMGDYQLLGAEEEHAATGSSPSSIANRVSYFLNLHGPSE
AVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSKAGMLSSDGRCKTFGKGADGYVPGEGVGAVL
LKPLSKAFADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGIDPRTISYIEAHGTGTSLGDPIEINGL
KKAFRQLKQFCAIGSVKSNIGHLESAAGIAGLIKVLLQMKHKTLVPSLHSEELNPYIDFEDSPFYVQQELQEWKRPVE
LPRRAGVSSFGAGGVNAHVVLEEYIPPAAGGPALIVLSAKNEERLREYARRLLDFLER-
DADLADIAYTLQVGREAMEERLAFVASSLEELEEKLEAFLDGEK-
VFRGQVKRNMDEAIDAWLEKGKLAKLAELWVKGAEIDWNRLYGPRRISLPTYPFARERYWVPEKESKAA--HE

Sector 1　Sector 2　Sector 3 Sector 4　Sector 5　Sector 6　Sector 7
　　Sector 8　　Sector 9　　Sector 10　Sector 11　Sector 12

Clustal-Omega — KR–PP-binding–KS — Conservation-filtered

VCYRNGKRYVAYKEIP-SGSVPFKDDGVYLITGGTGGLGLLFARHLAERYGAKLVLTGRSPLPPKEKI-
ELEAKGVQVEYLSADLTDREAVERLIEEIKRTFGPIGGVIHCAGVADDSFIRKTLEDIQEVLEPKVAGLQNLDEATKDE
PLDFFVLFSSVSAVIGNAGQSDYAMANAFMDYFAAYNRQGKTLSINWPLWKEGGMGEEVEEWLKELFSEELKIPAE
QLDTDEPFQDYGVDSILLAQLAQRLNKKFG-
LDPTVFFEYPTIESLAEYLLTEYPSAEDIAIIGMSCRFPGADDLEEFWDNLREGRDAITEVPAERWDWRSSKWGGFID
GIDQFDPLFFNISPREAELMDPQERLFLEESWKALEDAGYTREELSGKKVGVFVGGRSGHYPLLGALESKNPIVGGG
QNYIAARVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECFAAVAGGVNLLIHPSKYLMLSQAGMLSPDGRCHT
FDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPTAPNPEAQAEVIKSALEKAGIDPETISY
IEAHGTGTSLGDPIEIKALTKVFRSDDKQFCGIGSVKSNIGHLESAAGIASLIKVLLQLKHKQLPPSLHCEEPNPYIDFEK
SPFYVNRELEEWKRAPRRAGISSFGFGGTNAHVIVEEYEESAKLSAKILKEYAI

Sector 1　Sector 2　Sector 3 Sector 4　Sector 5　Sector 6　Sector 7
　　Sector 8　　Sector 9　　Sector 10　Sector 11　Sector 12

Figure S14 continued.

MAFFT with ep = 0.0 — KS–adapter domain–KR — Conservation-filtered

PKEPEPEALAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPGKTYCKWGGFLDGIDEFDPLFFN
ISPREAELMDPQQRLFLQEAWKALEDAGYTPESLGRTGVFVGVMNNGYQLLGAATGNSFSIIANRISYFLNLKGPSIP
VDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGMLSPDGRCKTFDAGADGYVPGEGVGAVL
LKPLSDAEADGDHIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYEKAGIDPETISYIEAHGTGTKLGDPIEINGLS
RAFRELTDDKFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHFETLNPHIDFDDSPFYVNTELQPWDG-
PRRAGVSSFGAGGTNAHIVIEYPAR-PE-
PALIVLSAKNEERLKEYAEQLLDFLRIDLADLAYTLQTGREAMEERLAFVAGSREELEEKLEAFLDG-
NEGIYRGQVKRNKDTLALFTADEDMAALVDAWIEKGAKLAELWVKGLDIDWNRLYPRRISLPTYPFAKERYWLP-
PED-
FKEGGVYLITGGTGGLGLLCARHLAERYGVKKLVLTGREPLPKVKAIQALEAKGVQVEYLSADLTDEAAVRQALE
QIKQTMGPIGGVIHCAGVVENPAFIRKTAEDIQRVLEPKVAGLQTLDEALANEPLDFFVLFSSVSAVIPAGQSDYAMA
NAFMDYFAIVSIQWPNWKETGMVRTKALKSGLLSLTNAEGLALLD

MAFFT with ep = 0.0 — PP-binding–KS — Conservation-filtered

ASSPAA-
EKVEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKFGELSPTLFFEYPTIAELAEYLLEEYPA-
AEDIAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDWREYSKGGFIDGIDEFDPLFFGISPREAELMDPQQRL
FLEEAWKAIEDAGYTPEELSKKVGVFVGVMNGDYQLAEA-ATG-
SQSIANRVSYFFNLHGPSEAVDTACSSSLVAIHLACQSLRNGECEMAIAGGVNLLLLHPEKYIMLSQAGMLSPDGRCKT
FDERADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTIS
YIEAHGTGTSLGDPIEINALTKAFRS-
DKCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKSPFYVQRELEEWKGPRRAGISSFGAG
GTNAHVILEEYEAR-PLSAKNERLREYA

MAFFT with ep = 0.0 — PP-binding–KS–adapter domain — Conservation-filtered

-E-
LREKVELKKLVSEILKIPAEDIDADTPLEEYGFDSISLTELTNRLNEEFGELSPTLFFEYQTIAELAAYLLEEHPAEDIAII
GISGRFPGADDLEEFWENLKEGKDCITEIPEERWDWREYYDPPGKTYSKWGGFIDGVDEFDPLFFGISPREAFLMDP
QERLFLETAWKAIEDAGYTPESLAKVGVFVGVMNGDYQLLGE-
ATGLSPSSIANRVSYFLNLIGPSEAVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSQAGMLSSDGR
CKTFGKGADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGIDP
RTISYIEAHGTGTSLGDPIEINGLKKAFRQLTDKFCAIGSVKSNIGHLESAAGIAGLTKVLLQMKHKTLVPSLHSEELNP
YIDFEDSPFYVQQELQEWKRG-PRRAGVSSFGAGGVNAHVVLEEYPPA--GPALIVLSAKNEERLREYARRLLDFLEQ-
DLADIAYTLQVGREAMLERLAFVASSLEELEEKLEAFLDGEKEGVFRGQVKRNADEDMDETIEAWLEKGKLAKLAE
LWVKGAEIDWNRLYKPRRISLPTYPFAREFRYWVPHPILHENT

Figure S14 continued.

MAFFT with ep = 0.0 — KR–PP-binding–KS — Conservation-filtered

ASSPAA-
EKVEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKFGELSPTLFFEYPIIAELAEYLLEEYPA-
AEDIAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDWREYSKGGFIDGIDEFDPLFFGISPREAELMDPQQRL
FLEEAWKAIEDAGYTPEEISKKVGVFVGVMNGDYQLAEA-ATG-
SQSIANRVSYFFNLHGPSEAVDTACSSSLVAIHLACQSLRNGECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKT
FDERADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTIS
YIEAHGTGTSLGDPIEINALTKAFRS-
DKCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKSPFYVQRELEEWKGPRRAGISSFGAG
GTNAHVILEEYEAR-PLSAKNERLREYA

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6   Sector 7
        Sector 8   Sector 9

MAFFT with ep = 0.123 — KS–adapter domain–KR — Conservation-filtered

APKPASRKPRPAAKPEPKAAAADEEPIAIIGMSGRYPGAPDLDEFWDNLAEGKDCITEIPKDRWDWRAYYDPKKGKT
YCKWGGFLDGIDEFDPLFFNISPREAELMDPQQRLFLQEAWKALEDAGYTPESLSGRTGVFVGVMNNGYGLLGNEQ
AEHAATGNSPSIANRISYFLNLKGPSIPVDTACSSSLVAIHLACQALRNGECDMAIAGGVNLYLTPEKYISLSKAGMLS
PDGRCKTFDAGADGYVPGEGVGAVLLKPLSDAEADGDHIYGVIRGSAINHGGKTNGITAPNPKAQADLIRDAYEKA
GIDPETISYIEAHGTGTKLGDPIEINGLSRAFREYTDRKQFCAIGSVKSNIGHLEAAAGIAGLIKVLLQMKHKTLVPSLHF
ETLNPHIDFDDSPFYVNTELQPWERDGAPRRAGVSSFGAGGTNAHIVIFEYQPKARAPEEPALIVLSAKNEERLKEYAE
QLLDFLRANIDLADLAYTLQTGREAMEERLAFVAGSREELEEKLEAFLDGKTNEGIYRGQVKRNKDTIALFEADEDM
AALLDAWIRKGKLAKLAELWVKGLDIDWNRLY-GEKPRRISLPTYPFAKERYWLP-
PEDPFKEGGVYLITGGTGGLGLLCARHFAERYGVKKLVLTGREPLPPRKEAAIRELEAKGVQVEYLSADLTDEAAV
RQALEQIKRTMGPIGGVIHCAGVEDPAFIRKTAEDIQRVLEPKVAGLQTLDEALANEPLDFFVLFSSVSAVIGSAGQS
DYAMANAFMDYFAAIVSIQWPNWKETGMEVETKAYQEGLLSLTNAEGLQLLDR

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6   Sector 7
        Sector 8   Sector 9   Sector 10   Sector 11   Sector 12

MAFFT with ep = 0.123 — PP-binding–KS — Conservation-filtered

--K-AAAPAQEEAPAAESDADL-
EKVEEWLKELFSEELKIPAEQIDTDTPFEEYGFDSILLAQLANRLNKKFGELSPTLFFEYPTIAELAEYLLEEYPAAAE
DLAIIGMSGRFPGADDLEEFWENLKEGKDCITEIPAERWDGGKTYSKWGGFIDGIDEFDPLFFGISPREAELMDPQQR
LFLEEAWKAIEDAGYTPEELSGKKVGVFVGVMHGDYQLLL-AGNPIAATGN--
SIANRVSYFFNLHGPSEAVDTACSSSLVAIHLACQSLRNGECEMAIAGGVNLLLHPEKYIMLSQAGMLSPDGRCKTFD
ERADGYVPGEGVGAVLLKPLSKAEADGDHIYAVIKGSAVNHGGRTNGLTAPNPKAQADLIKKALEKAGIDPRTISYI
EAHGTGTSLGDPIEINALTKAFR-
LTDDKQCGIGSVKSNIGHLESAAGIAGLIKVLLQLKHKTLPPSLHCEELNPYIDFEKSPFYVQRELEEWPDGPRRAGISS
FGAGGTNAHVILEEY--A-A--RRKP-LIVLSAKNEERKEYA

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6   Sector 7
        Sector 8   Sector 9

Figure S14 continued.

MAFFT with ep = 0.123 — PP-binding–KS–adapter domain — Conservation-filtered

--AES-ADA--
AASEEEEASLREKVEAYLKKLVSEVLKIPAEEIDADTPLEEYGFDSISLTELTNRLNEEFGELSPTLFFEYQTIAELAAY
LLEEH-AADDLAIIGISGRFPGADDLEEFWENLKEGKDCITEIPEDRWDWRAYYDPD-
PGKTYSKWGGFIDGVDEFDPLFFGISPREAFLMDPQERLFLETAWKAIEDAGYTPESLATKVGVFVGVMNGDYQLL
GEEG-
AATGNYSSIANRVSYFLNLHGPSEAVDTACSSSLVAIHLACESLRRGECEMAIAGGVNLILHPNKYISLSQAGMLSSDG
RCKTFGKGADGYVPGEGVGAVLLKPLSKAEADGDHIYGVIKGSAVNHGGKTNGLTVPNPNAQADLIREALRKAGID
PRTISYIEAHGTGTSLGDPIENGLKKAFRQLTDKQFCAIGSVKSNIGHLESAAGIAGLIKVLLQMKHKTLVPSLHSEEL
NPYIDFEDSPFYVQQELQEWKRNPRRAGVSSFGAGGVNAHVVLEEYIPPAAAA-
DGPALIVLSAKNEERLREYARRLLDFLERE--TDLADIAYTLQVGRFAMEERLAFVASSLEELEEKLEAFLDGEK-
EGVFRGQVKRNKEALA--DE-DELIEAWLEKGKLAKLAELWVKGAEIDWNRLYGRKPRRISLPTYPFARFRYWVP-----
-RLHPLLHENTS

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6   Sector 7
          Sector 8   Sector 9   Sector 10   Sector 11

MAFFT with ep = 0.123 — KR–PP-binding–KS — Conservation-filtered

VCYRNGKRYVAYLKEIP---S-
VPFKDDGVYLITGGTGGLGLLFARHLAERGVAKLVLTGRSPLTSKKEKIKELEAKGVQVEYLSADLTDREAVERLIE
EIKRTFGPIGGVIHCAGVADDPAFIRKTLEDIQEVLEPKVAGLQNLDEATKDEPLDFFVLFSSVSAVIGNAGQSDYAMA
NAFMDYFAAYRKGQGKTLSINWPLWKEGGMGEKVEEWLKELFSEELKIPAEQLDTDEPFQDYGVDSILLAQLAQRL
NKKFGDLDPTVFFEYPTIESLAEYLLSEYPASEEIAIIGMSCRFPGADDLEEFWDNLREGRDAITEVPAERWDWSNSK
WGGFIDGIDQFDPLFFNISPREAELMDPQARLFLEESWKALEDAGYTREELSGKKVGVFVGGRSGHYPDEERGNPIV
GNGQNIAARVSYFFNLRGPSIVLDTACSSSLVAIHLACQSLRSGECEAAVAGGVNLLLHPSKYLMLSQAGMLSPDGR
CHTFDERADGYVPGEGVGAVLLKPLSQAIADGDRIYAVIKGSAVNNDGRTNGPTAPNPEAQAEVIKSALEKAGIDPE
TISYIEAHGTGTSLGDPIELKALTKVFTDDKQFCGIGSVKSNIGHLESAAGIASLIKVLLQLKHKQLPPSLHCEEPNPYID
FEKSPFYVNRELEEWEPRRAGISSFGFGGTNAHVIVEEYEDEQ--KRTPLPPELAKTEERLDEYAE

Sector 1   Sector 2   Sector 3   Sector 4   Sector 5   Sector 6   Sector 7
          Sector 8   Sector 9   Sector 10   Sector 11   Sector 12

Figure S14 continued.

Figure S15. Comparison of RoseTTAFold and Phyre2 models. Predicted protein structures for the KS-adapter-KR (adapter referring to the flanking subdomain) tridomain using the RosettaFold server (A-C) and the Phyre2 server (D-F). The protein sectors obtained from SCA and the NAHVILEE and LPTYPFx$_5$YW motifs are indicated by the colored sticks and the colored parts of the ribbon representation, respectively.
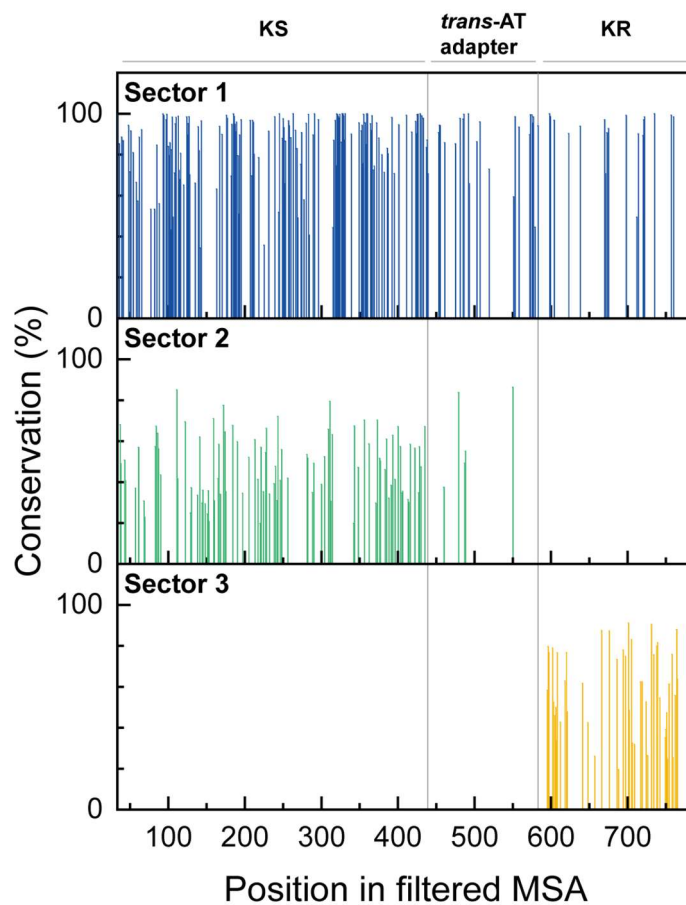
Figure S16. Positional conservation of the amino acid positions in the filtered MSA for sectors 1, 2 and 3 as obtained from the SCA of the MUSCLE alignments of the KS–*trans*-AT adapter–KR tridomain. The consensus sequence can be found in Figure S14.
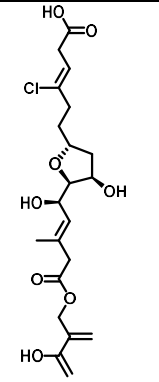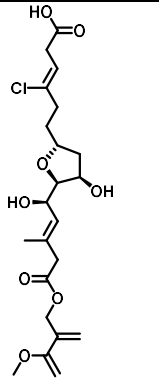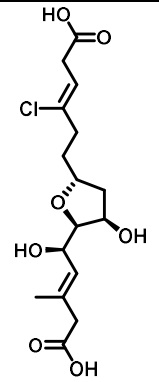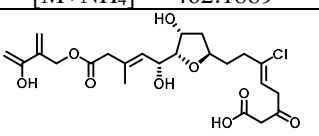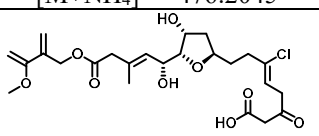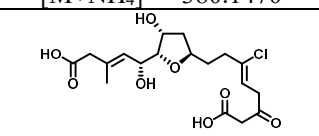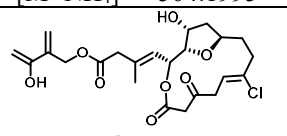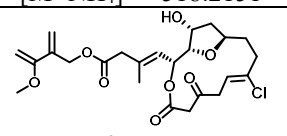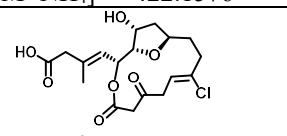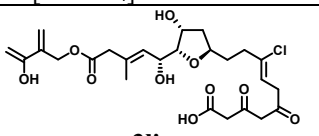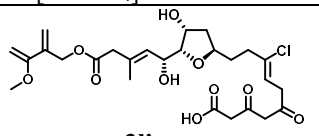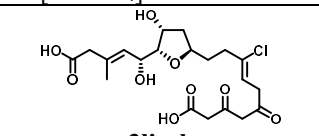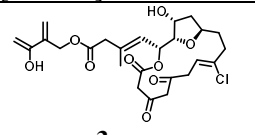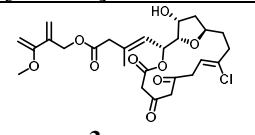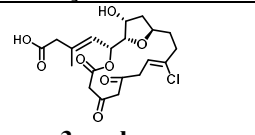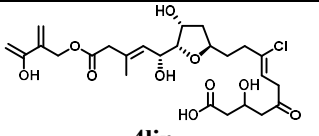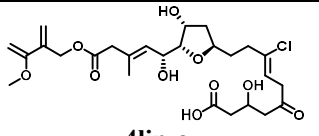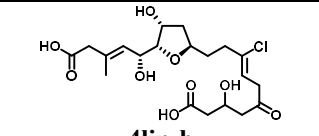
1
$C_{21}H_{29}ClO_8$
$[M+H]^+ = 445.1624$
$[M+NH_4]^+ = 462.1889$

1a
$C_{22}H_{31}ClO_8$
$[M+H]^+ = 459.1780$
$[M+NH_4]^+ = 476.2045$

1b
$C_{16}H_{23}ClO_7$
$[M+H]^+ = 363.1205$
$[M+NH_4]^+ = 380.1470$

2lin
$C_{23}H_{31}ClO_9$
$[M+H]^+ = 487.1730$
$[M+NH_4]^+ = 504.1995$

2lin-a
$C_{24}H_{33}ClO_9$
$[M+H]^+ = 501.1886$
$[M+NH_4]^+ = 518.2151$

2lin-b
$C_{18}H_{25}ClO_8$
$[M+H]^+ = 405.1311$
$[M+NH_4]^+ = 422.1576$

2cyc
$C_{23}H_{29}ClO_8$
$[M+H]^+ = 469.1624$
$[M+NH_4]^+ = 486.1889$

2cyc-a
$C_{24}H_{31}ClO_8$
$[M+H]^+ = 483.1780$
$[M+NH_4]^+ = 500.2045$

2cyc-b
$C_{18}H_{23}ClO_7$
$[M+H]^+ = 387.1205$
$[M+NH_4]^+ = 404.1470$

3lin
$C_{25}H_{33}ClO_{10}$
$[M+H]^+ = 529.1835$
$[M+NH_4]^+ = 546.2100$

3lin-a
$C_{26}H_{35}ClO_{10}$
$[M+H]^+ = 543.1992$
$[M+NH_4]^+ = 560.2257$

3lin-b
$C_{20}H_{27}ClO_9$
$[M+H]^+ = 447.1417$
$[M+NH_4]^+ = 464.1682$

3cyc
$C_{25}H_{31}ClO_9$
$[M+H]^+ = 511.1730$
$[M+NH_4]^+ = 528.1995$

3cyc-a
$C_{26}H_{33}ClO_9$
$[M+H]^+ = 525.1886$
$[M+NH_4]^+ = 542.2151$

3cyc-b
$C_{20}H_{25}ClO_8$
$[M+H]^+ = 429.1311$
$[M+NH_4]^+ = 446.1576$

4lin
$C_{25}H_{35}ClO_{10}$
$[M+H]^+ = 531.1992$
$[M+NH_4]^+ = 548.2257$

4lin-a
$C_{26}H_{37}ClO_{10}$
$[M+H]^+ = 545.2148$
$[M+NH_4]^+ = 562.2413$

4lin-b
$C_{20}H_{29}ClO_9$
$[M+H]^+ = 449.1573$
$[M+NH_4]^+ = 466.1838$

Figure S17. Predicted structures, chemical formulas, and masses as searched for in UHPLC-HRMS analysis.
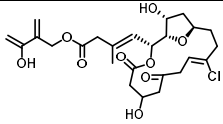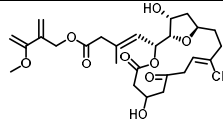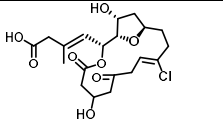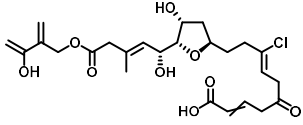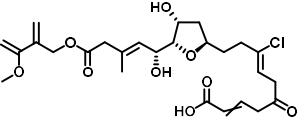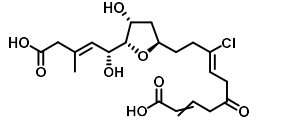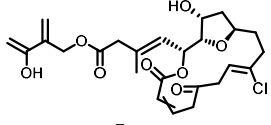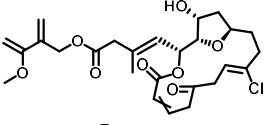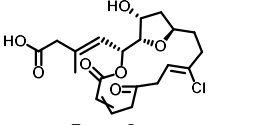
4cyc $C_{25}H_{33}ClO_9$ $[M+H]^+ = 513.1886$ $[M+NH_4]^+ = 530.2151$	**4cyc-a** $C_{26}H_{35}ClO_9$ $[M+H]^+ = 527.2042$ $[M+NH_4]^+ = 544.2308$	**4cyc-b** $C_{20}H_{27}ClO_8$ $[M+H]^+ = 431.1467$ $[M+NH_4]^+ = 448.1732$
5lin $C_{25}H_{33}ClO_9$ $[M+H]^+ = 513.1886$ $[M+NH_4]^+ = 530.2151$	**5lin-a** $C_{26}H_{35}ClO_9$ $[M+H]^+ = 527.2043$ $[M+NH_4]^+ = 544.2308$	**5lin-b** $C_{20}H_{27}ClO_8$ $[M+H]^+ = 431.1467$ $[M+NH_4]^+ = 448.1732$
5cyc $C_{25}H_{31}ClO_8$ $[M+H]^+ = 495.1780$ $[M+NH_4]^+ = 512.2046$	**5cyc-a** $C_{26}H_{33}ClO_8$ $[M+H]^+ = 509.1937$ $[M+NH_4]^+ = 526.2202$	**5cyc-b** $C_{20}H_{25}ClO_7$ $[M+H]^+ = 413.1362$ $[M+NH_4]^+ = 430.1627$

Figure S17 continued.
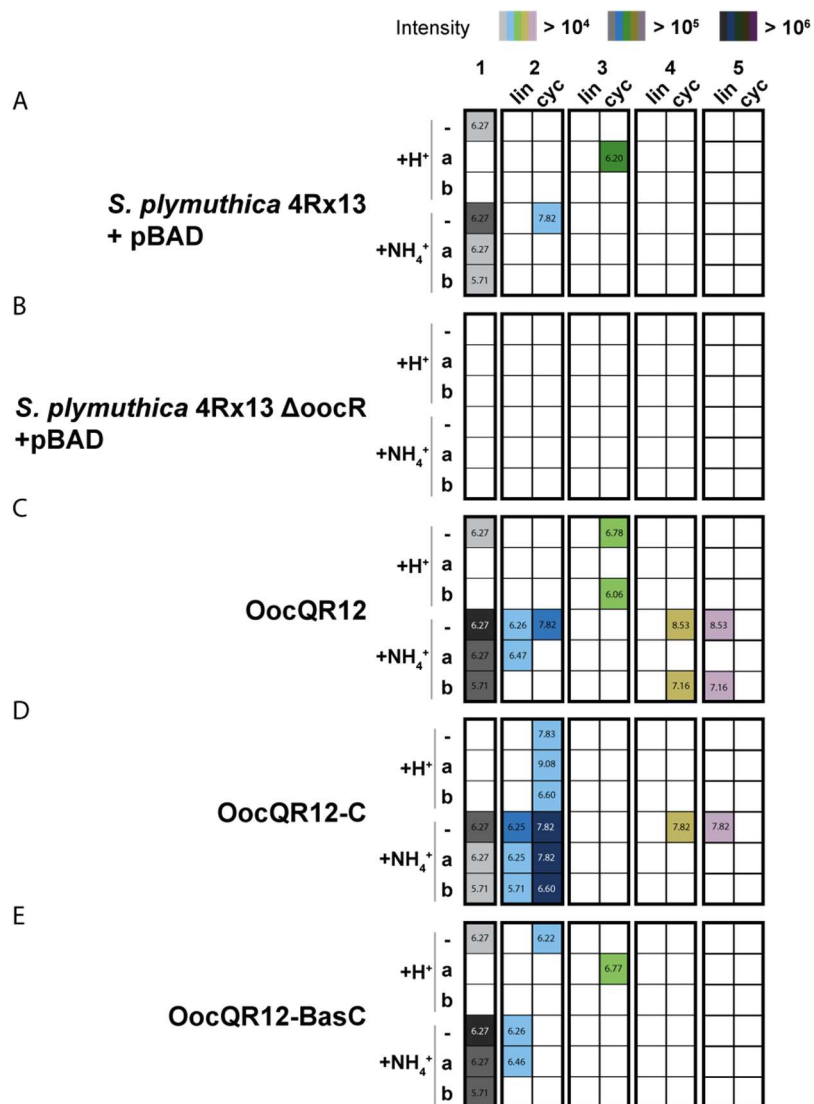
Figure S18. Heatmaps showing the presence, intensity range of the proton and ammonium adducts of the various compounds **1-5**, **1-5ₐ** and **1-5_b** in extract of cultures of A) wild type *S. plymuthica* 4Rx13 with empty pBAD vector, B) *S. plymuthica* 4Rx13 ΔoocR with empty pBAD vector, C) OocQR12, D) OocQR12-C and E) OocQR12-BasC grown under expression conditions. The retention times of the peaks are indicated in the heatmaps. Peaks were confirmed to have an isotope pattern indicating chlorination and in case multiple peaks were observed, the most intense peak is reported.
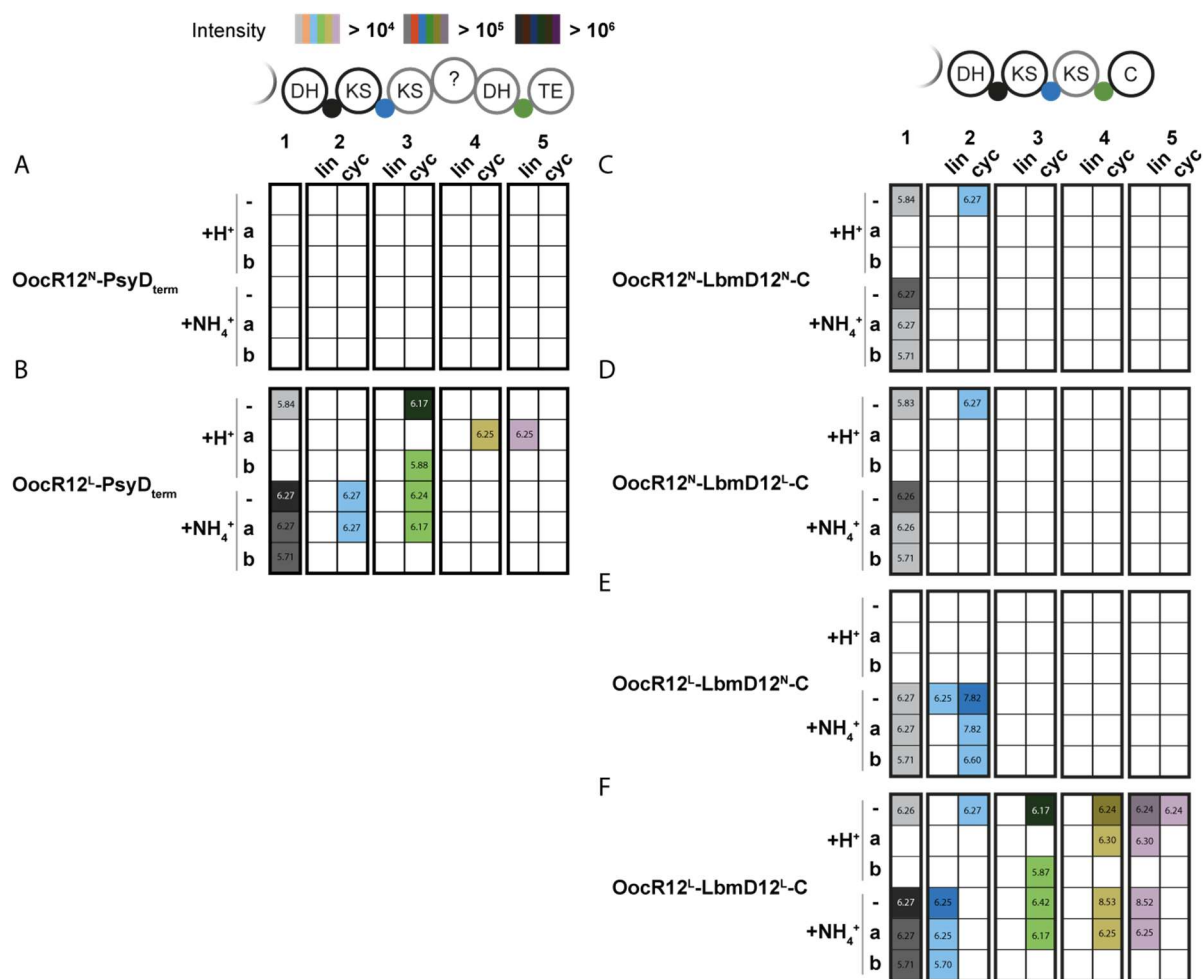
Figure S19. Heatmaps showing the presence, intensity range of the proton and ammonium adducts of the various compounds **1-5**, **1-5$_a$** and **1-5$_b$** in extract of cultures of A) OocR12^N-PsyD$_{term}$, B) OocR12^L-PsyD$_{term}$, C) OocR12^N-LbmD12^N-C, D) OocR12^N-LbmD12^L-C, E) OocR12^L-LbmD12^N-C and F) OocR12^L-LbmD12^L-C grown under expression conditions. The retention times of the peaks are indicated in the heatmaps. Peaks were confirmed to have an isotope pattern indicating chlorination and in case multiple peaks were observed, the most intense peak is reported.
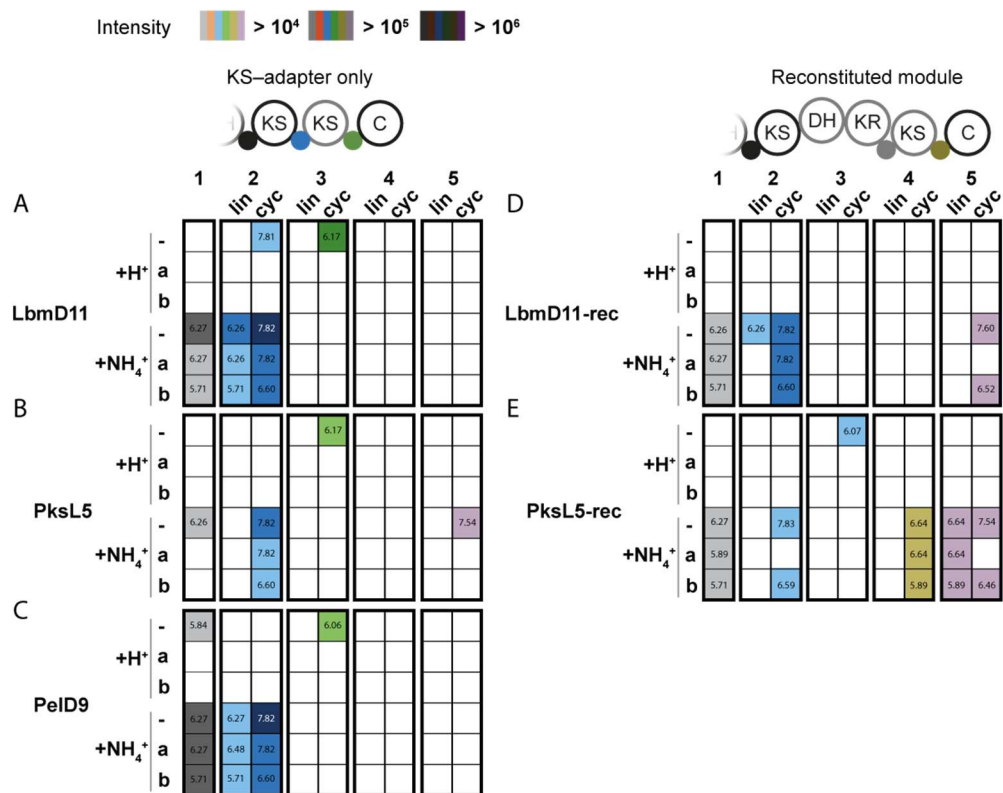
Figure S20. Heatmaps showing the presence, intensity range of the proton and ammonium adducts of the various compounds **1-5**, **1-5a** and **1-5b** in extract of cultures of A) LbmD11, B) PksL5, C) PelD9, D) LbmD11-rec and E) PksL5-rec grown under expression conditions. The retention times of the peaks are indicated in the heatmaps. Peaks were confirmed to have an isotope pattern indicating chlorination and in case multiple peaks were observed, the most intense peak is reported.
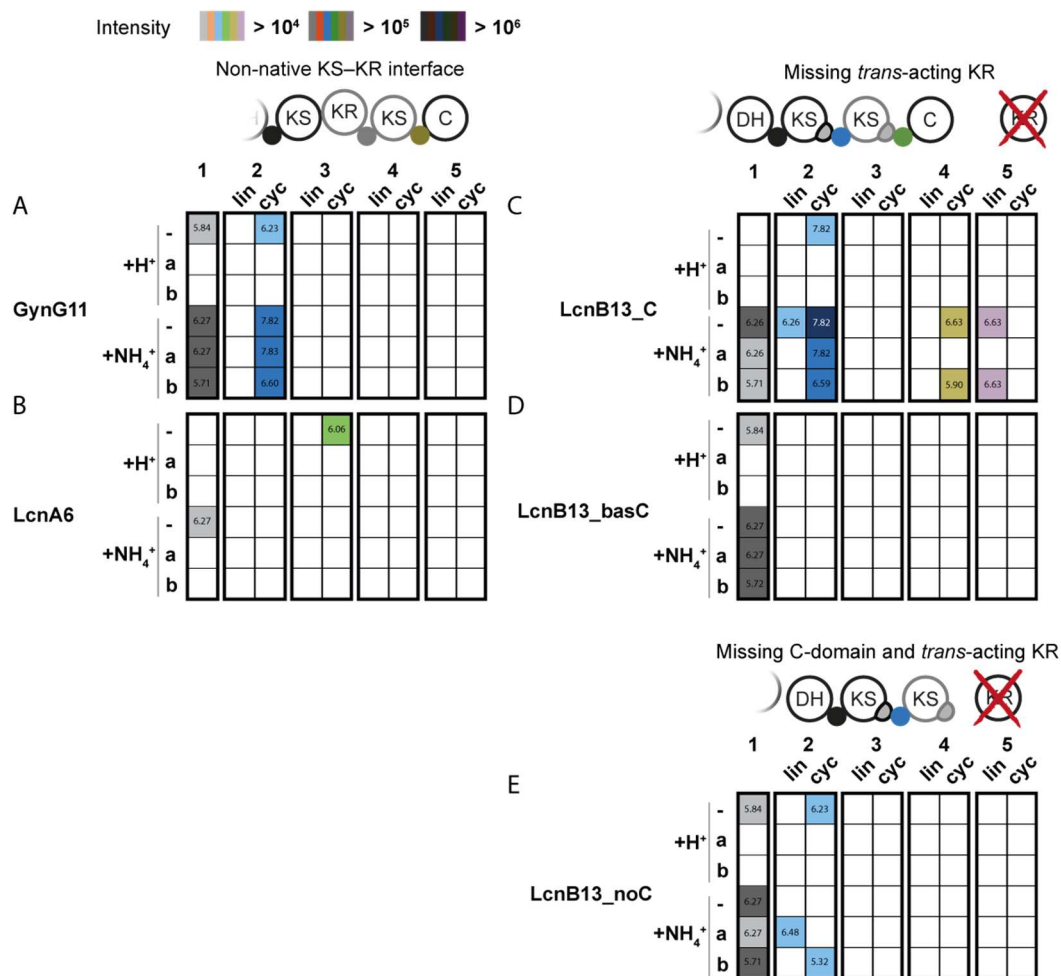
Figure S21. Heatmaps showing the presence, intensity range of the proton and ammonium adducts of the various compounds **1-5**, **1-5ₐ** and **1-5ᵦ** in extract of cultures of A) GynG11, B) LcnA6, C) LcnB13_C, D) LcnB13_basC and E) LcnB13_noC grown under expression conditions. The retention times of the peaks are indicated in the heatmaps. Peaks were confirmed to have an isotope pattern indicating chlorination and in case multiple peaks were observed, the most intense peak is reported.
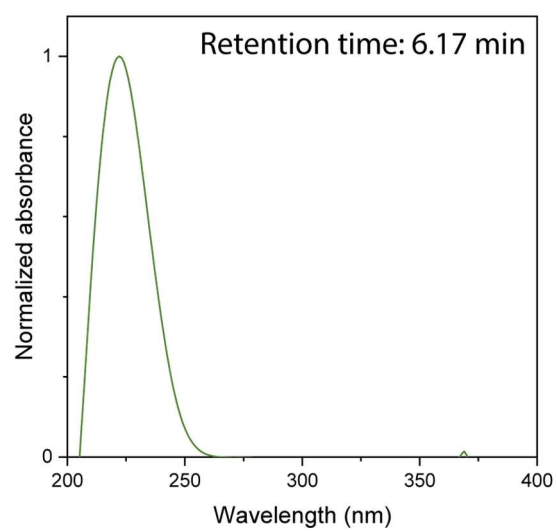
Figure S22. UV spectrum obtained from the UHPLC-MS analysis of extracts of OocR12^L-PsyD$_{term}$ of compounds with a retention time of 6.17 minutes, the time at which the mass corresponding to **3cyc** is observed.
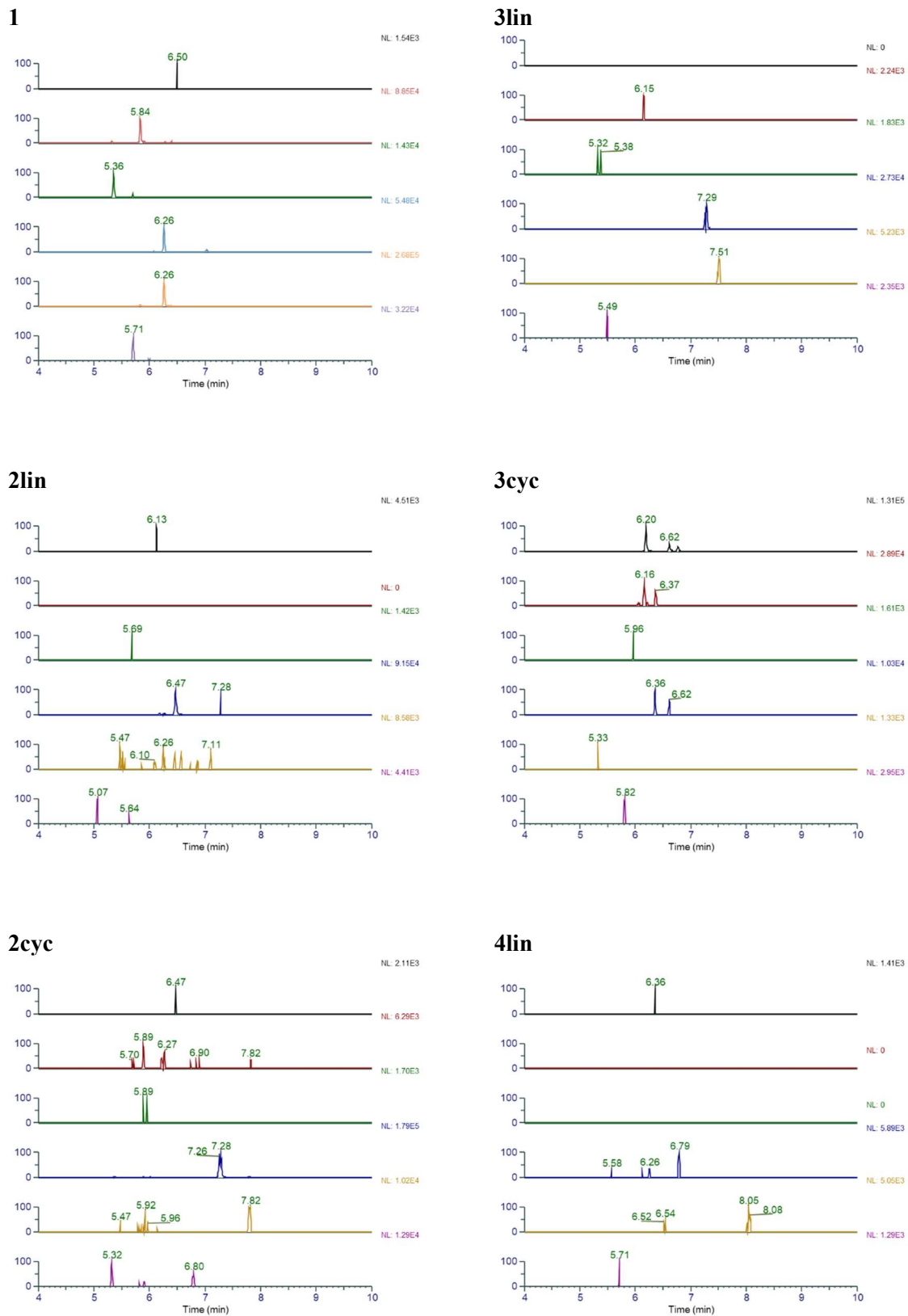
Figure S23. Extracted ion chromatograms for *S. plymuthica* 4Rx13 harboring the empty pBAD plasmid.

4cyc
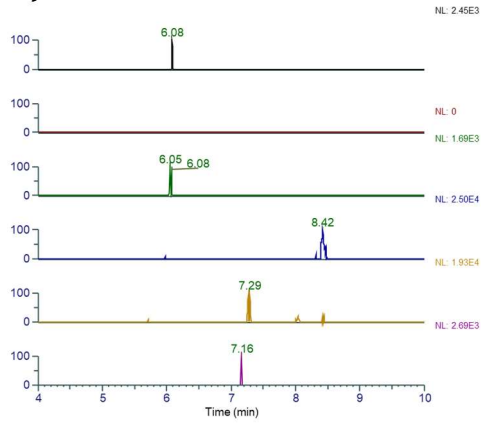
5lin
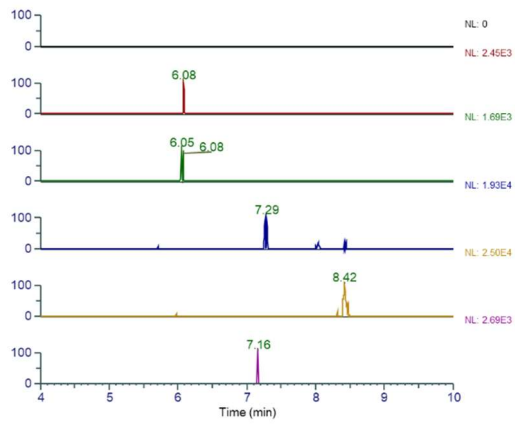
5cyc
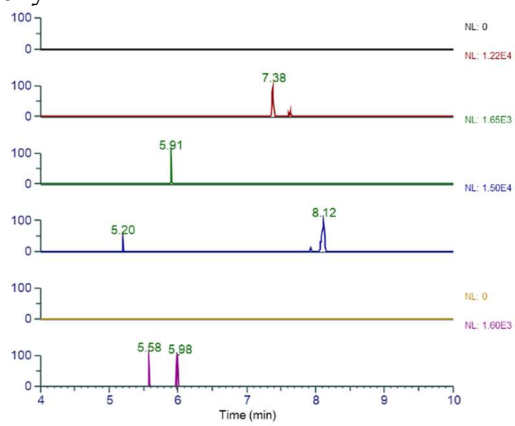
Figure S23 continued.
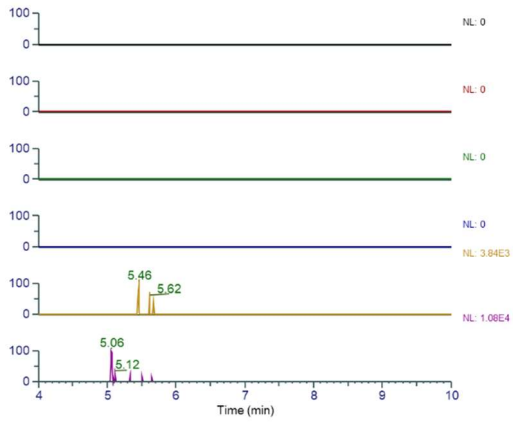
Figure S24. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring empty pBAD.

4cyc
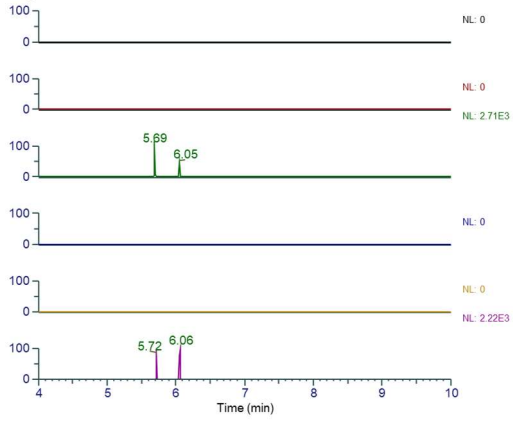
5lin
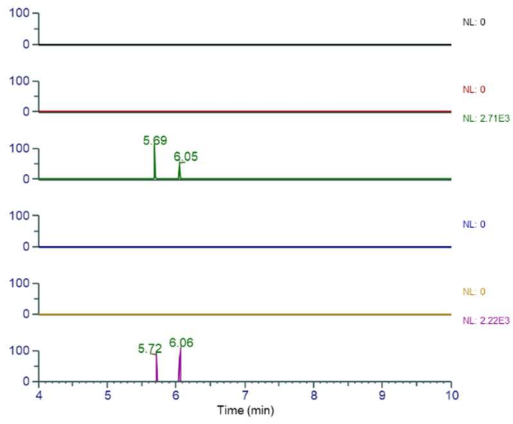
5cyc
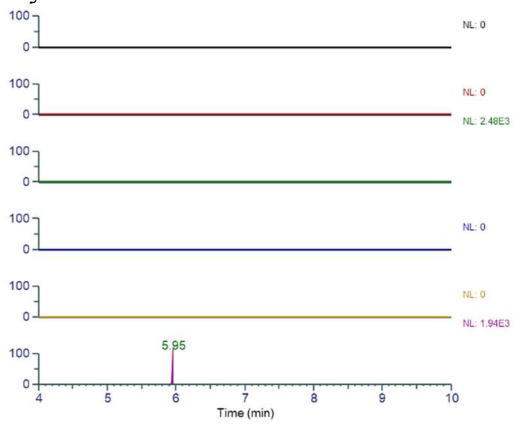
Figure S24 continued.
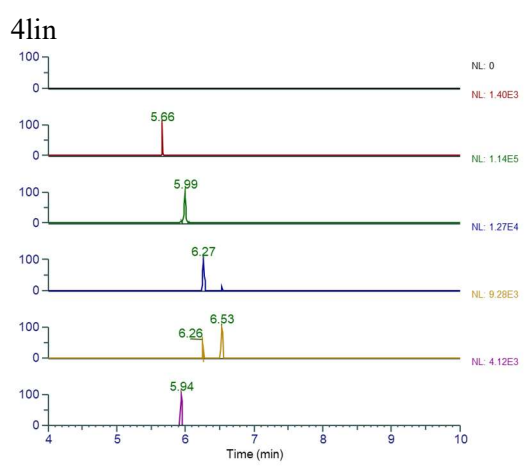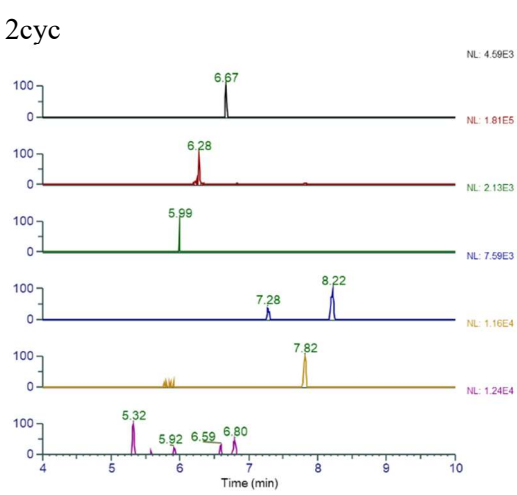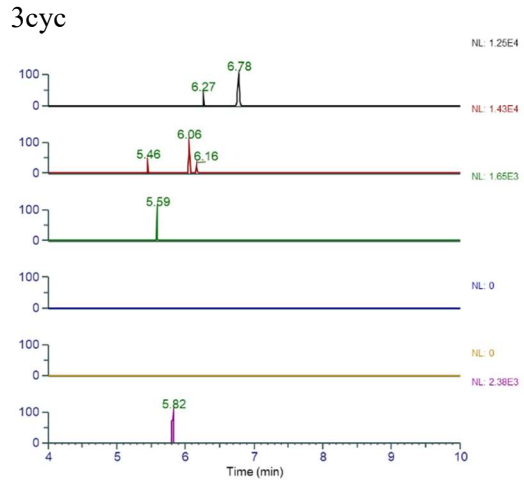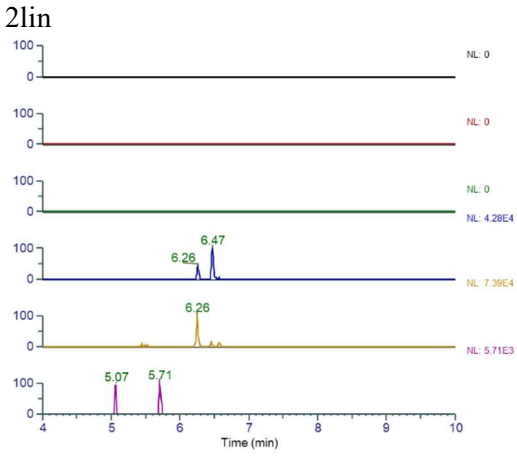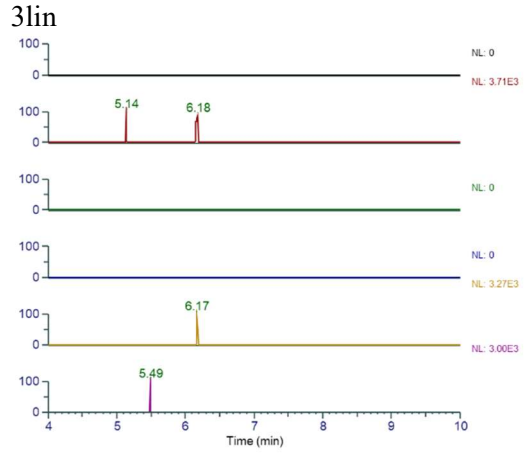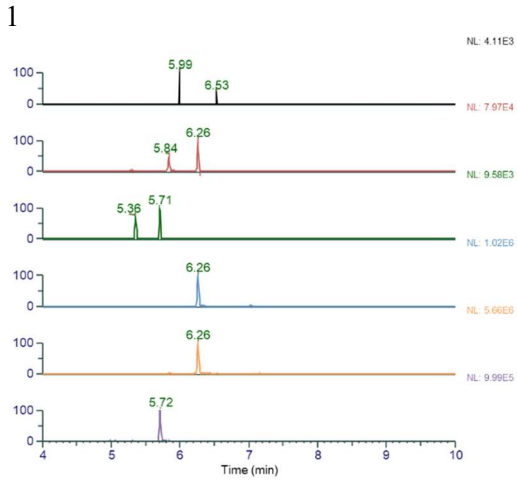
Figure S25. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR.

4cyc
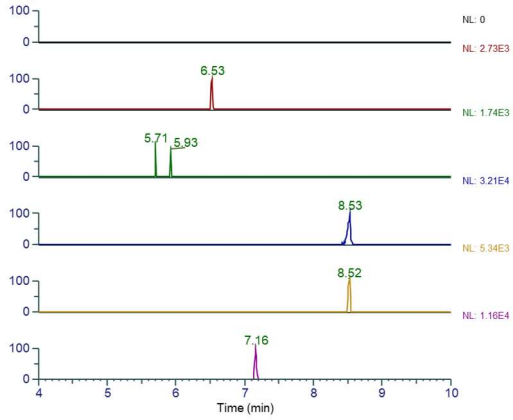
5lin
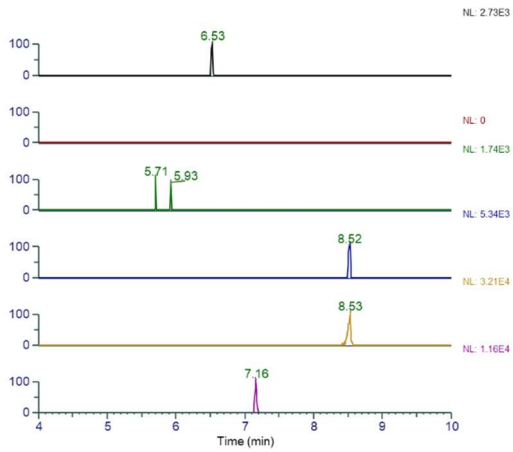
5cyc
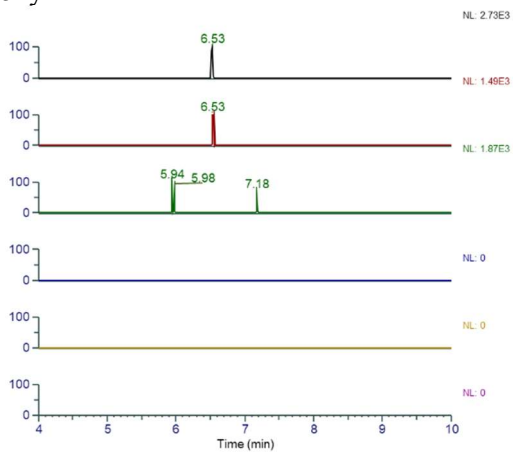
Figure S25 continued.
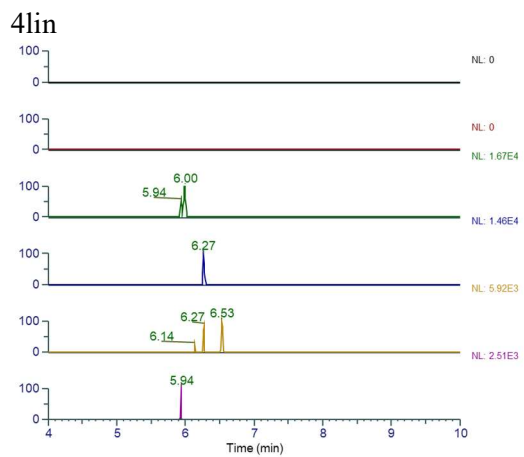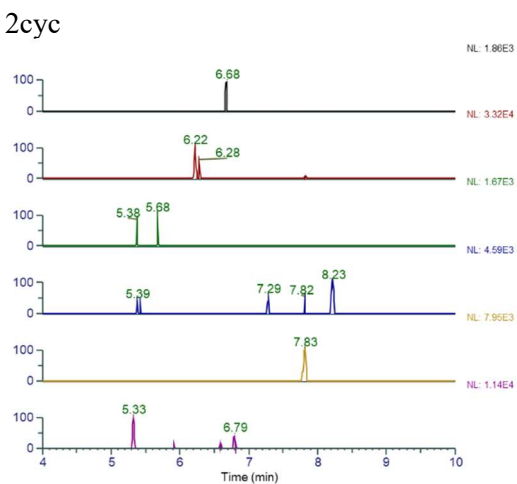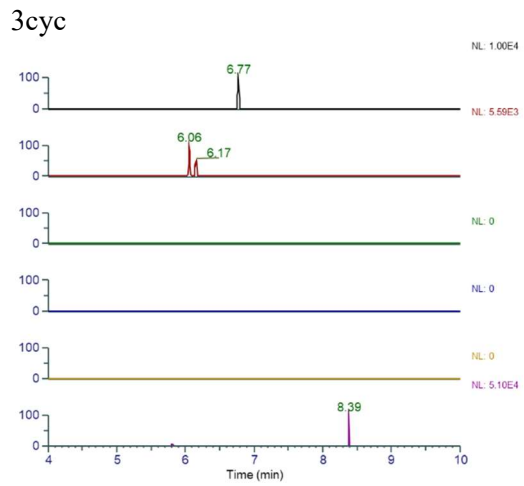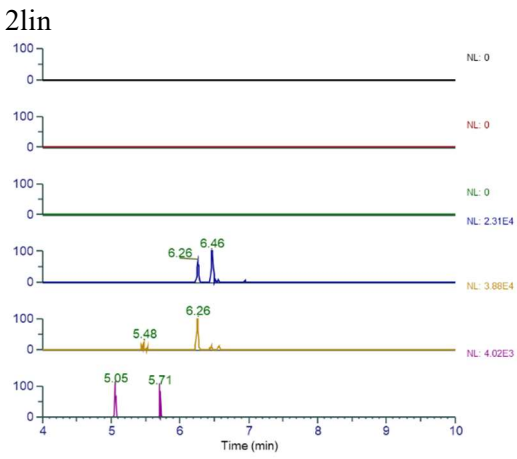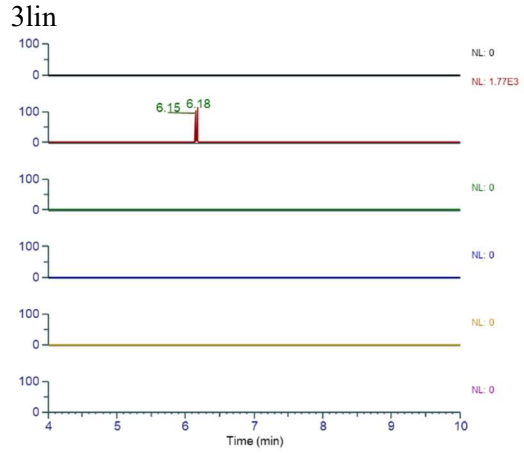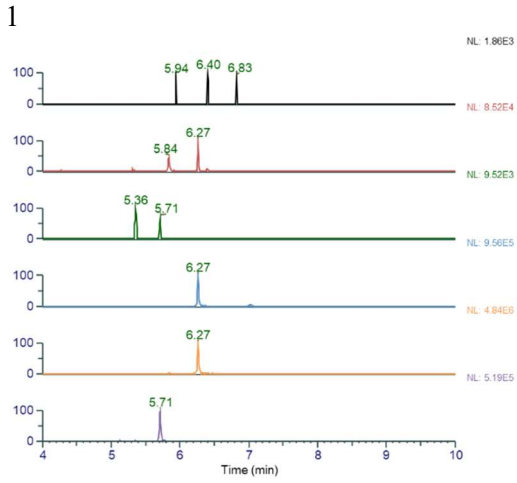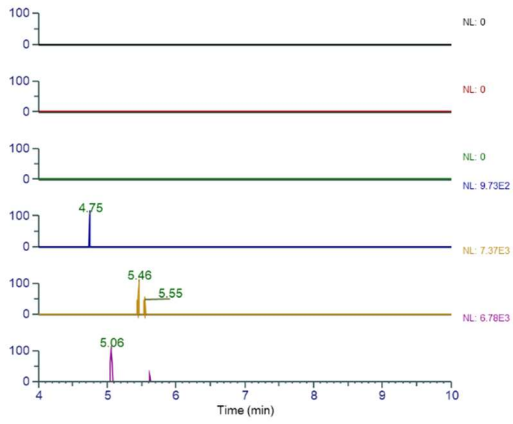
Figure S26. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-BasC.

4cyc
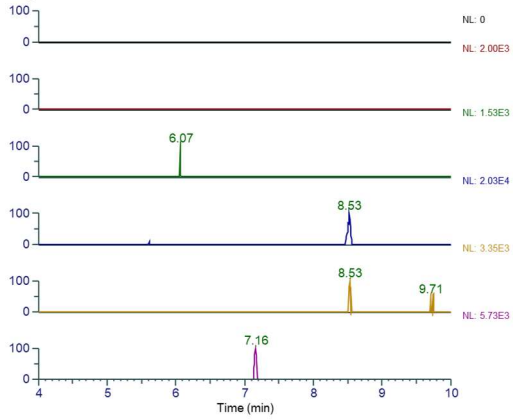
5lin
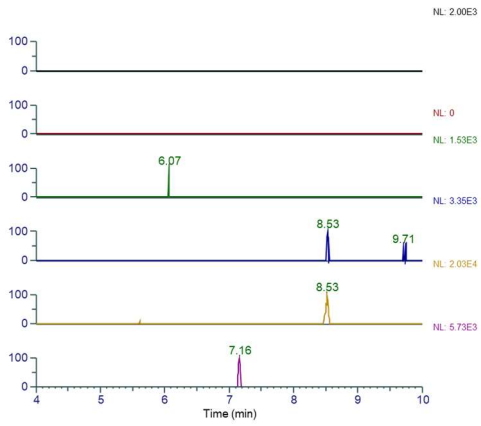
5cyc
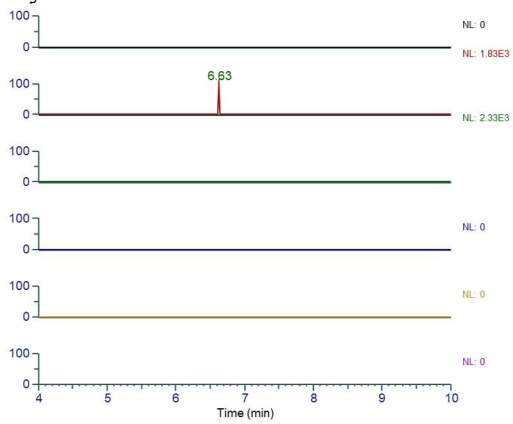
Figure S26 continued.
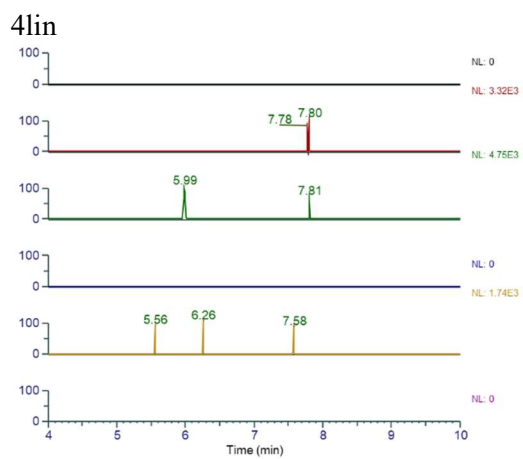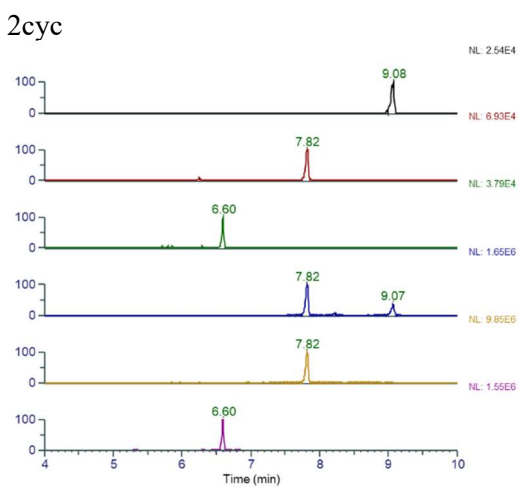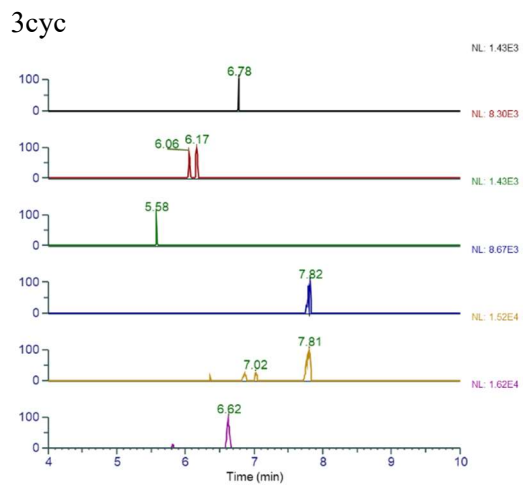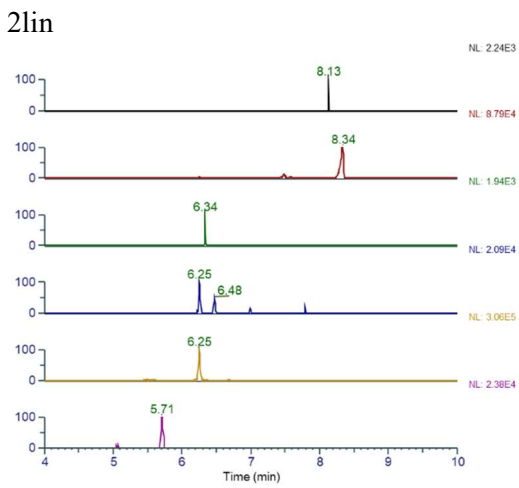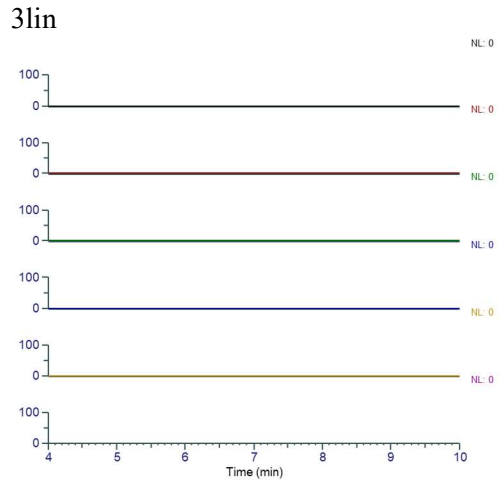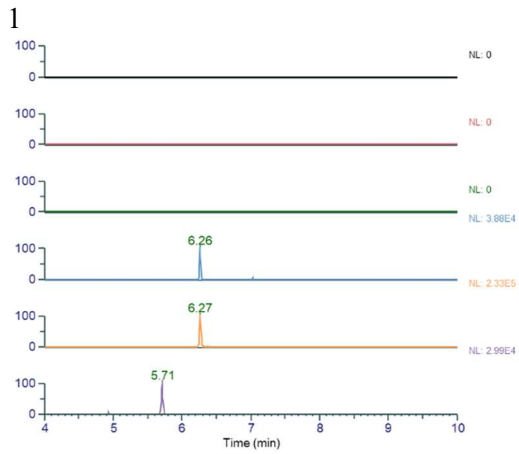
Figure S27. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQRC.

4cyc
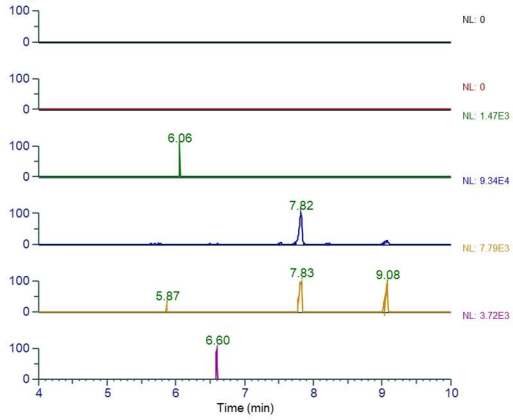
5lin
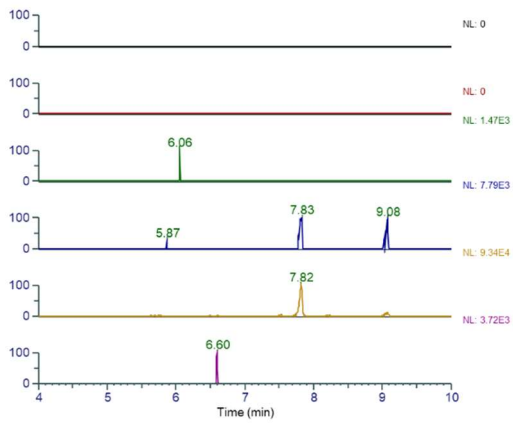
5cyc
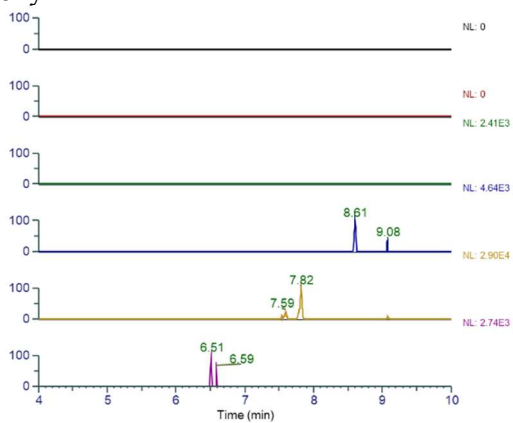
Figure S27 continued.
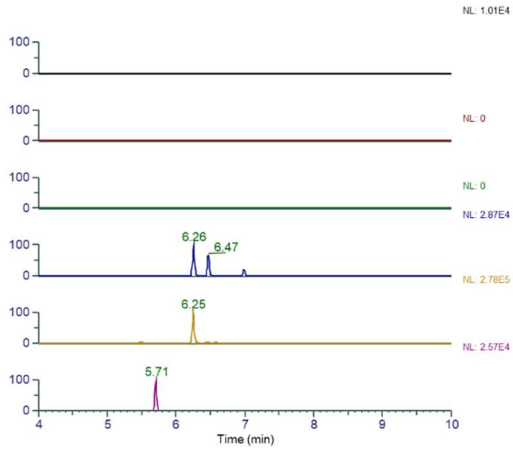
Figure S28. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-PsyD-L.

4cyc
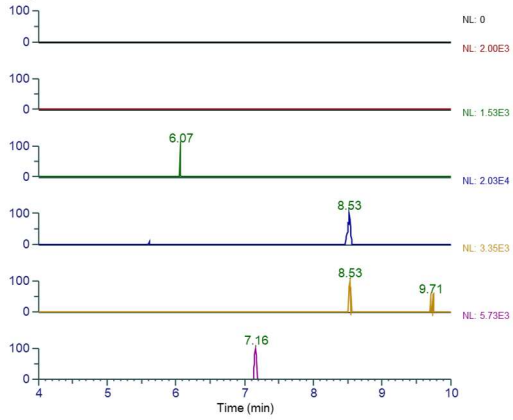
5lin
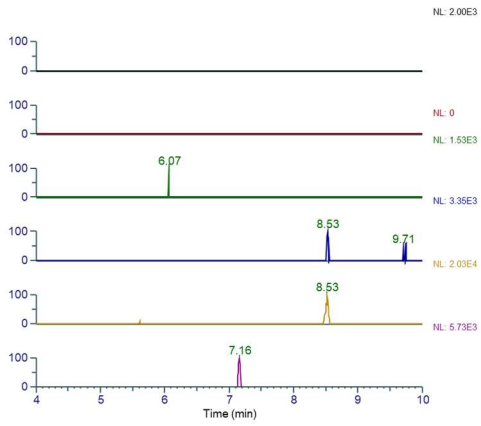
5cyc
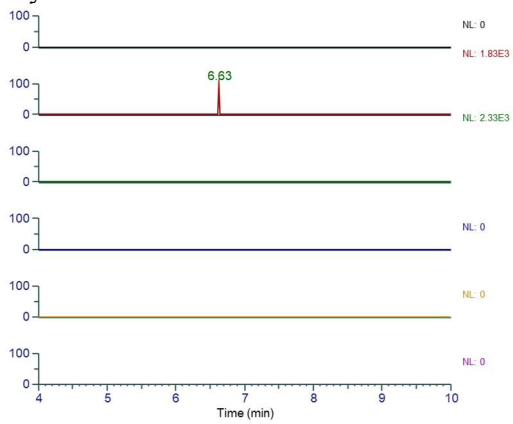
Figure S28 continued.

1
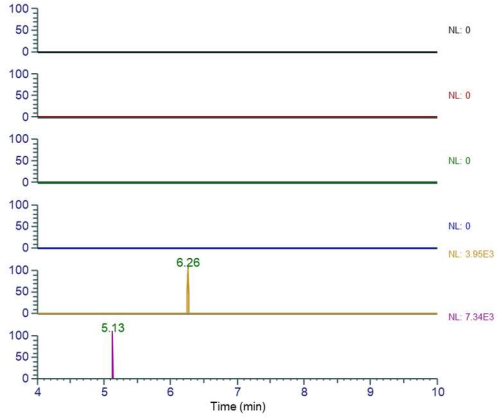
3lin
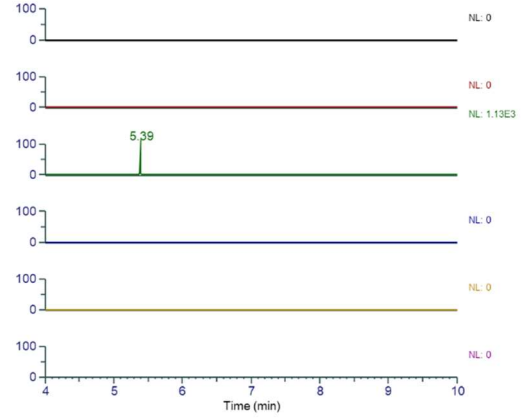
2lin
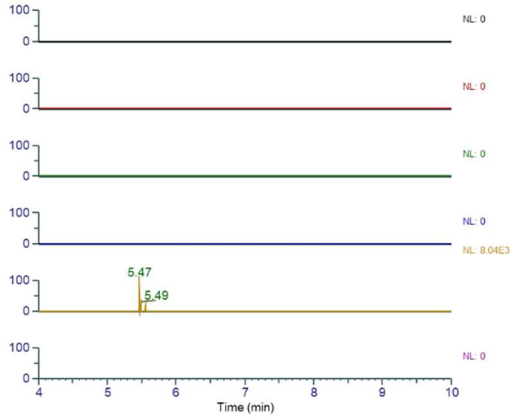
3cyc
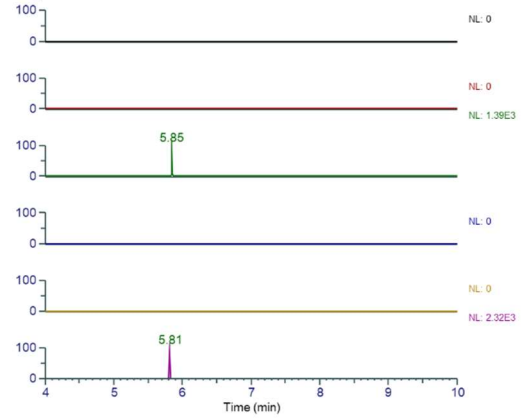
2cyc
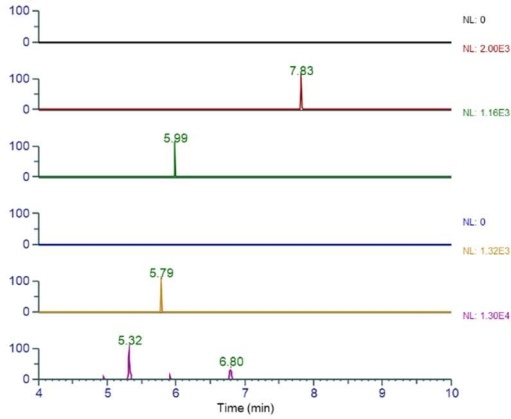
4lin
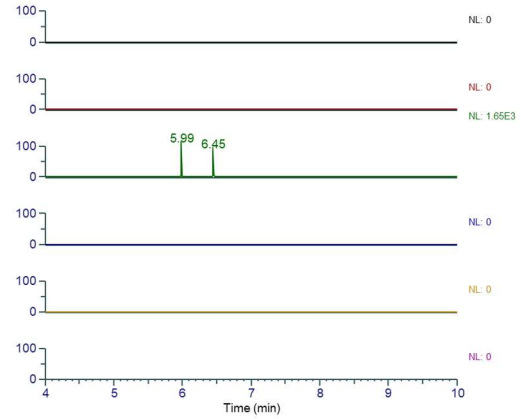
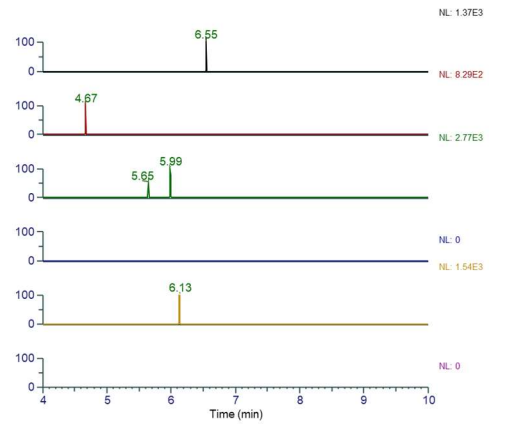**Figure S29.** Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-PsyD-N.

4cyc
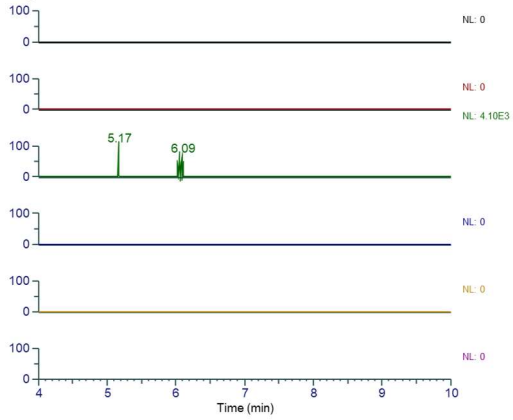
5lin
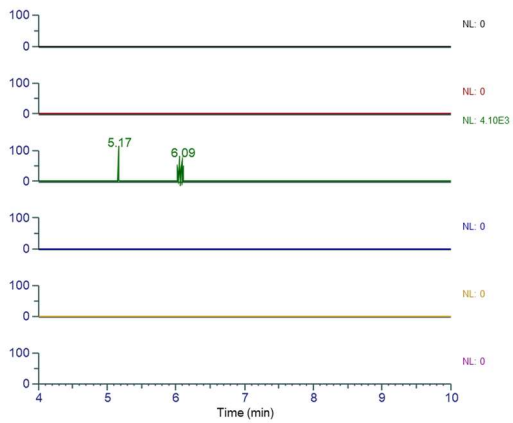
5cyc
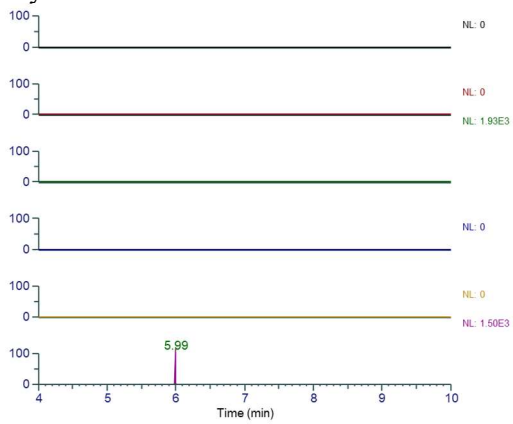
Figure S29 continued.
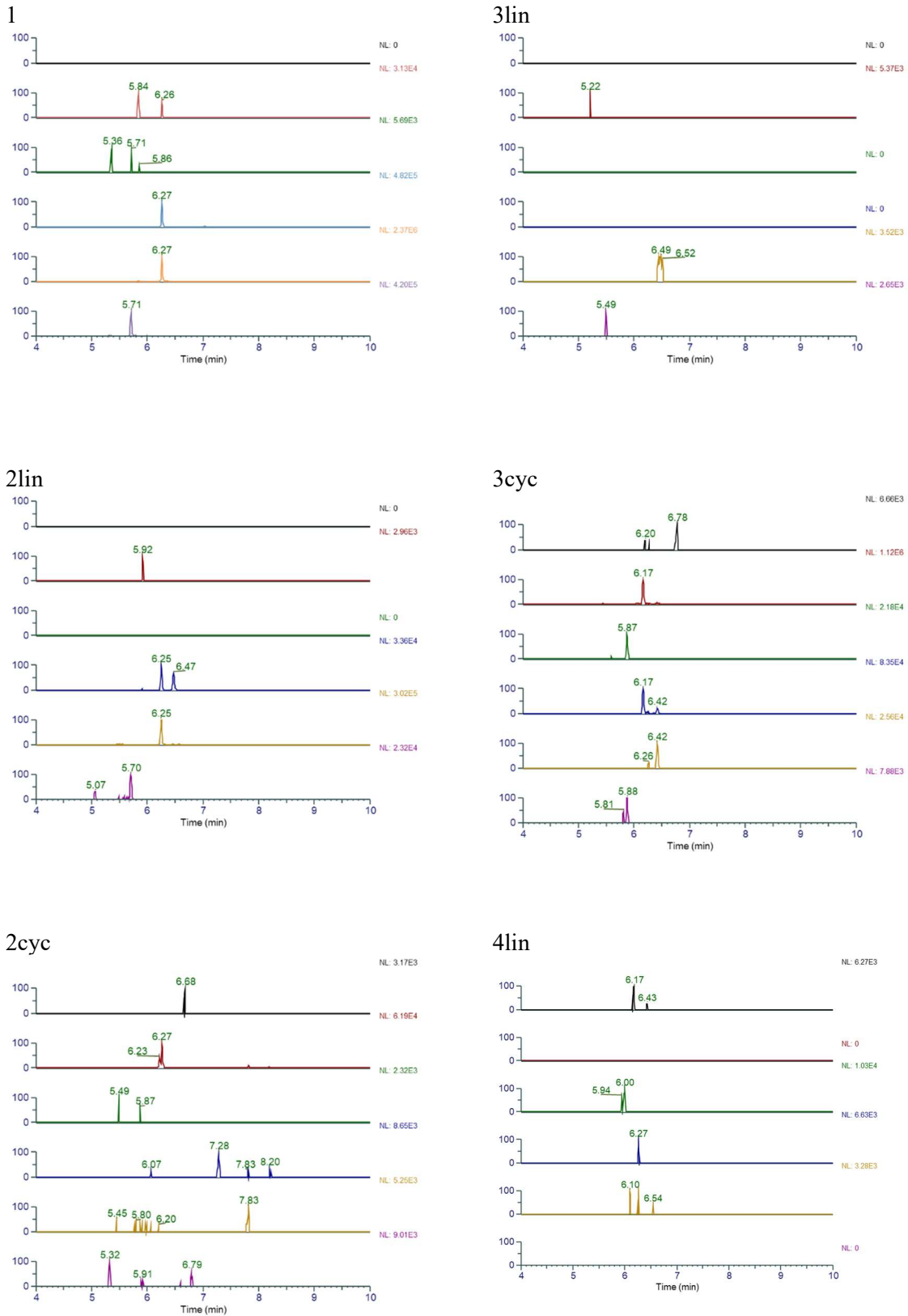
Figure S30. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-Lbm12-C LL.

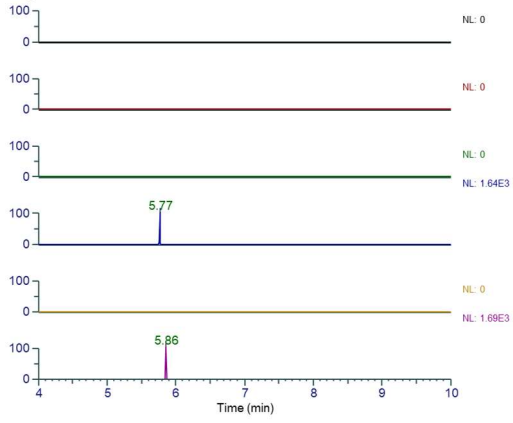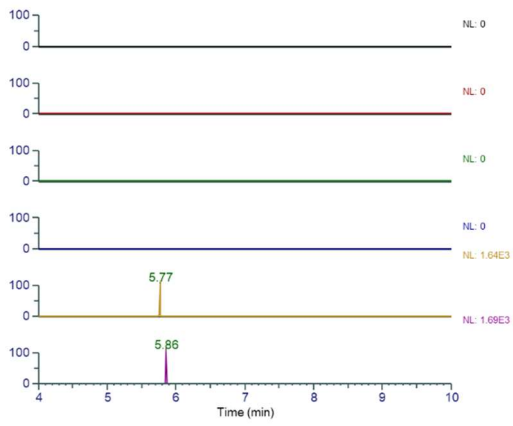4cyc
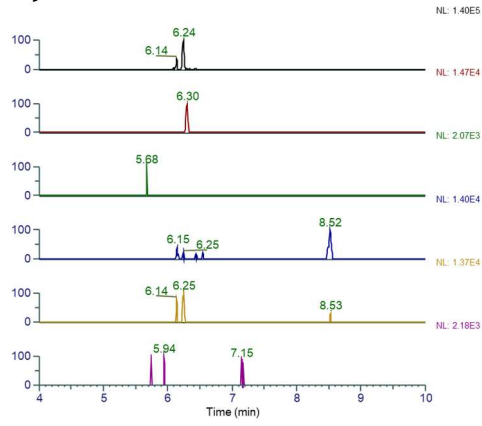
5lin
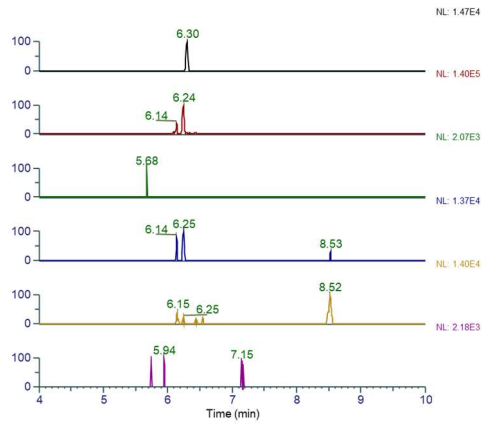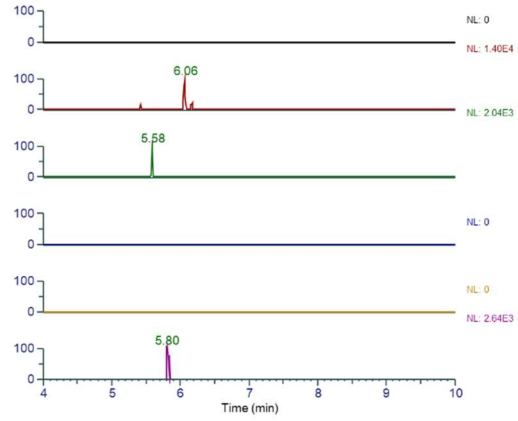
5cyc
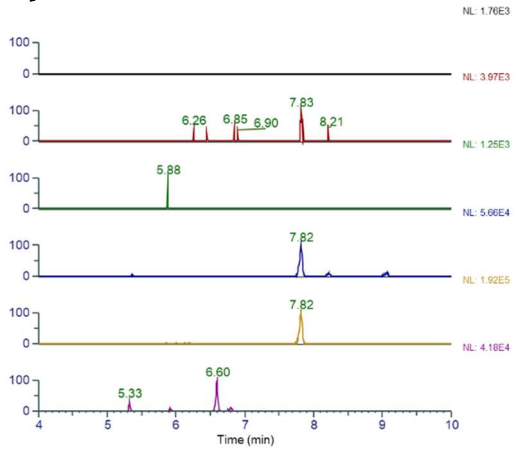
Figure S30 continued.
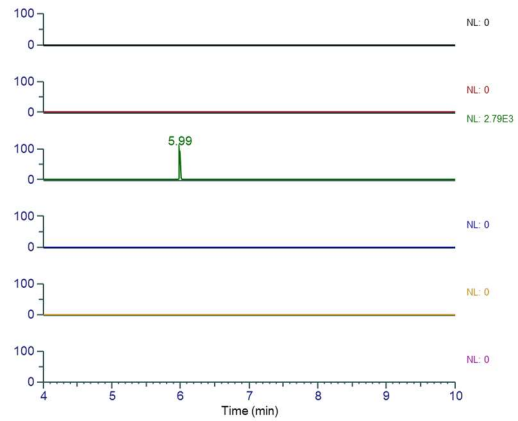
Figure S31. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-Lbm12-C LN.

4cyc
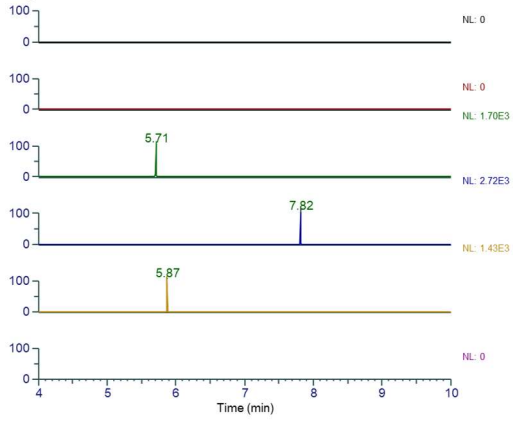
5lin
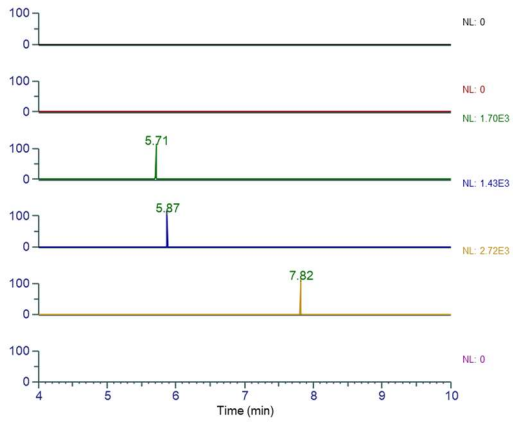
5cyc
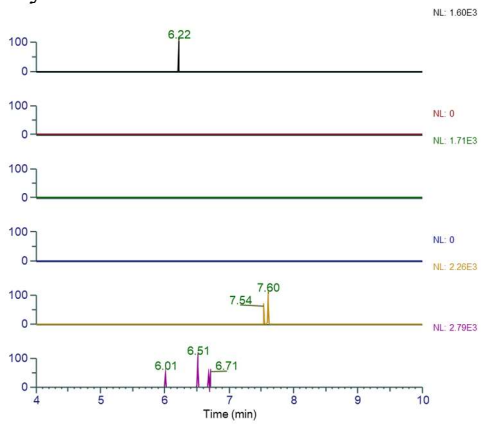
Figure S31 continued.
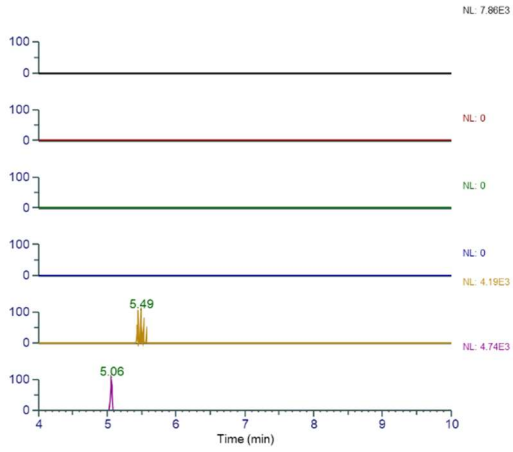
Figure S32. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-Lbm12-C NL.

4cyc
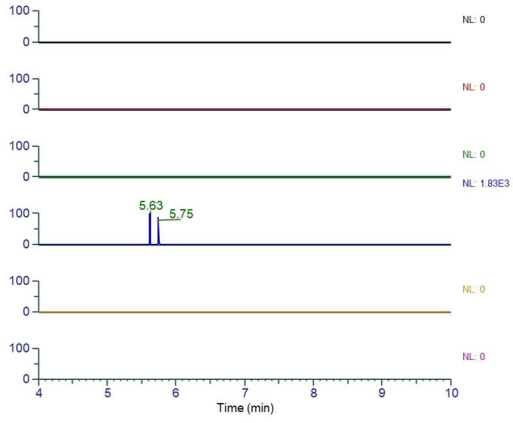
5lin
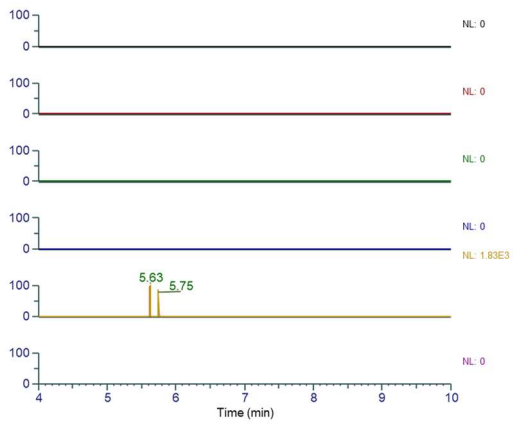
5cyc
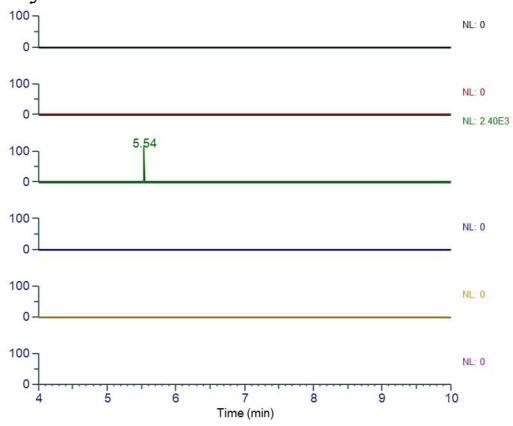
Figure S32 continued.
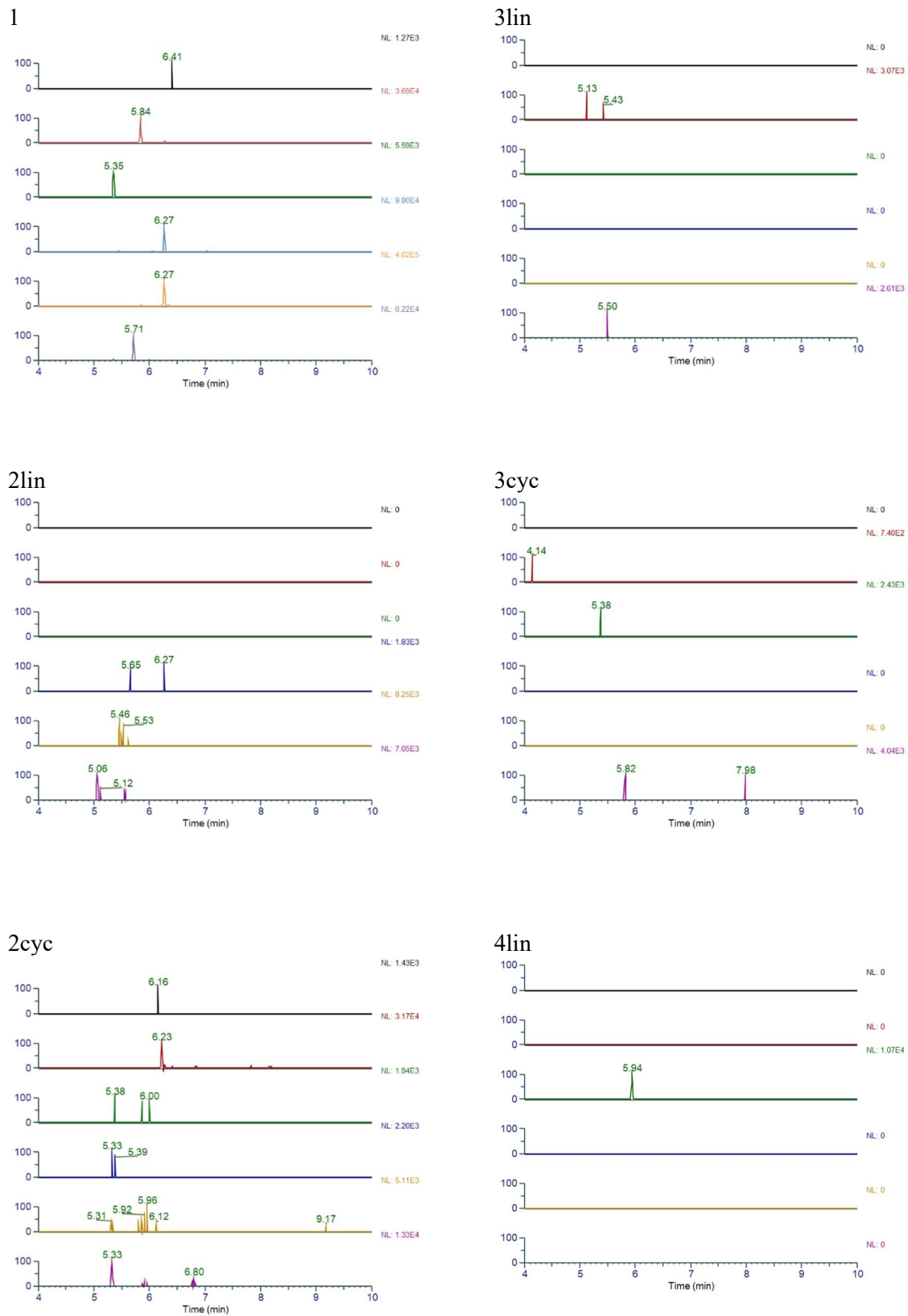
Figure S33. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-Lbm12-C NN.

4cyc
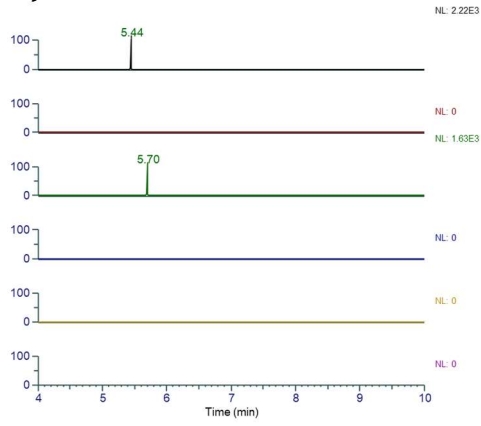
5lin
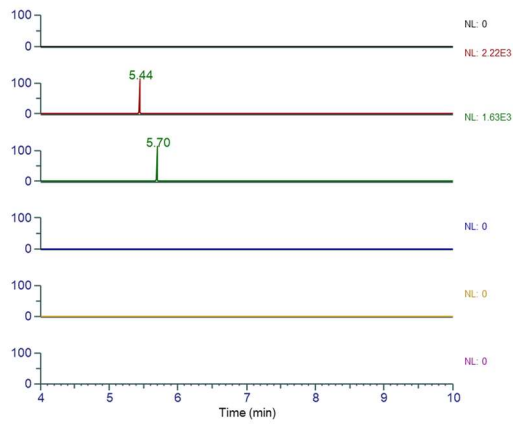
5cyc
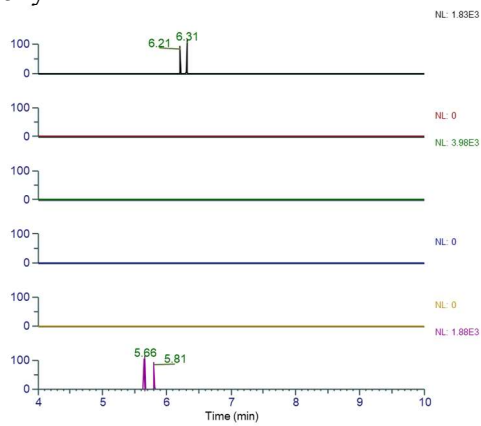
Figure S33 continued.
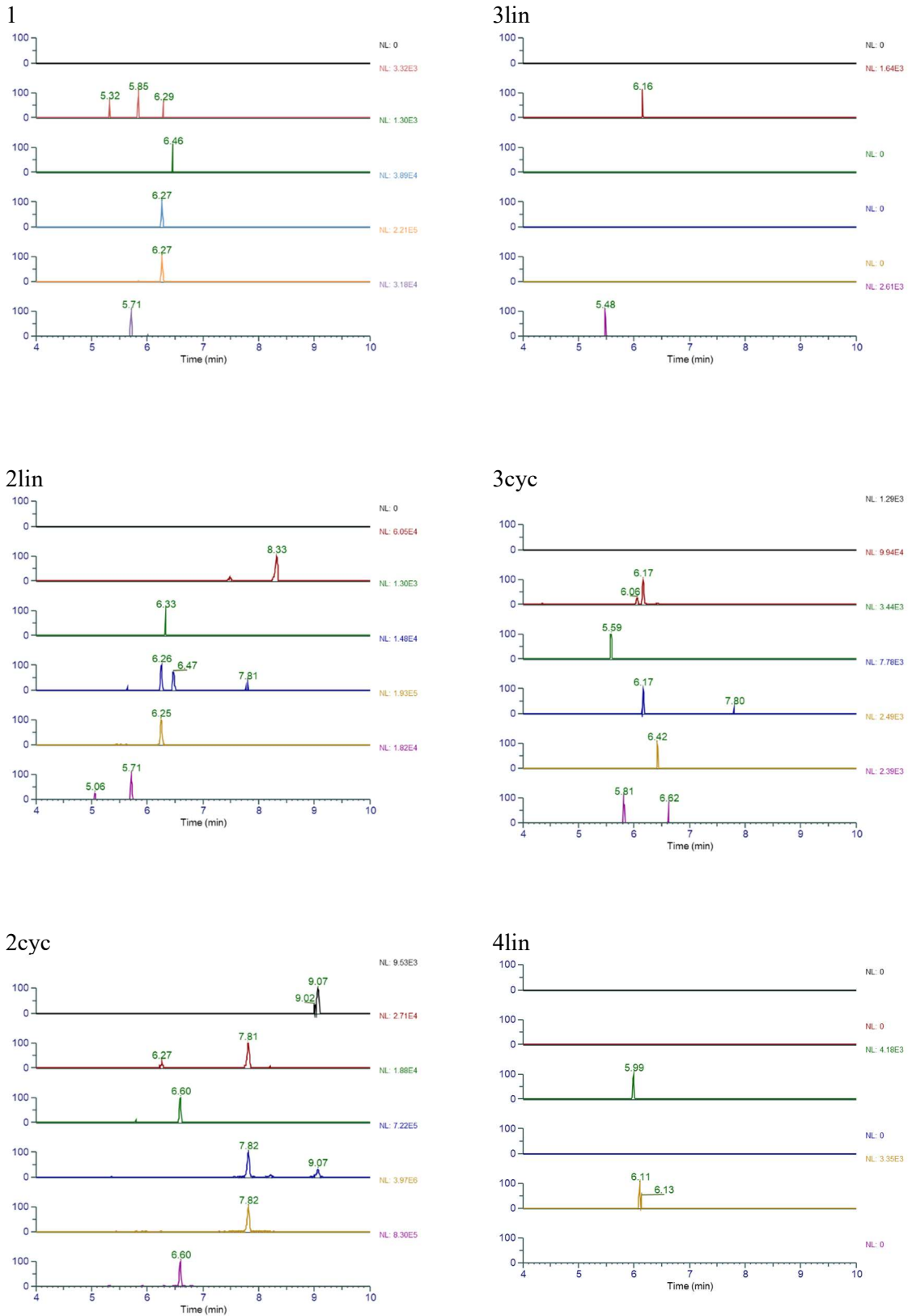
Figure S34. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-Lbm11-C.

4cyc
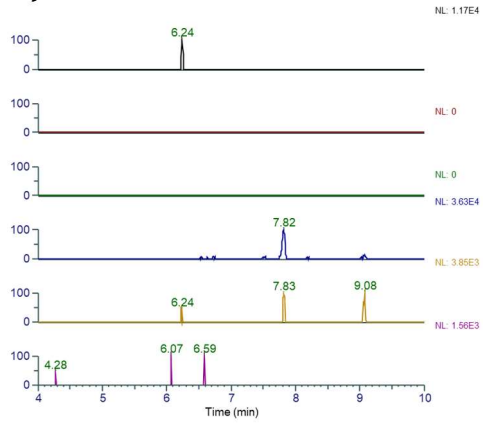
5lin
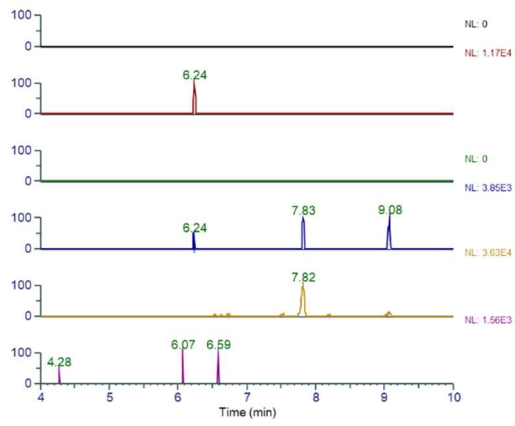
5cyc
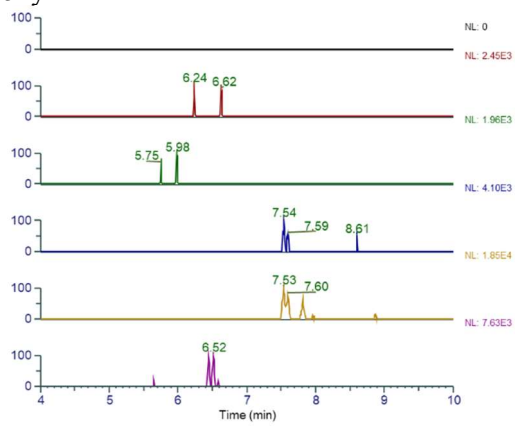
Figure S34 continued.
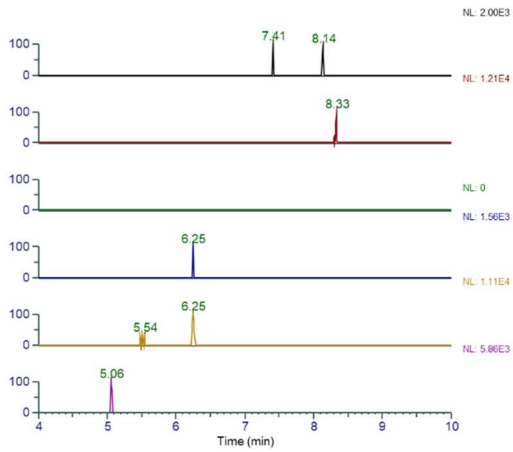
Figure S35. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-PksL5-C.

4cyc
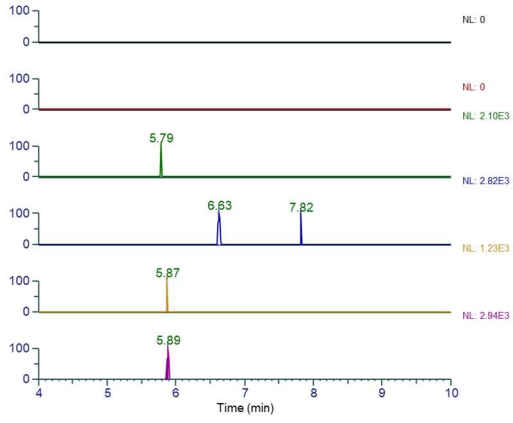
5lin
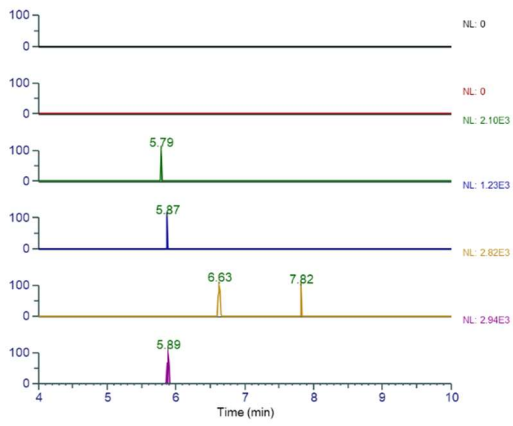
5cyc
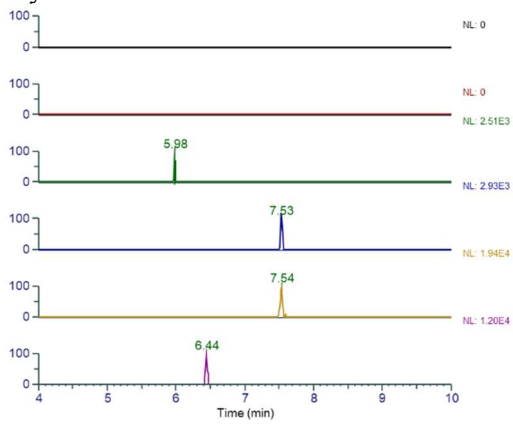
Figure S35 continued.
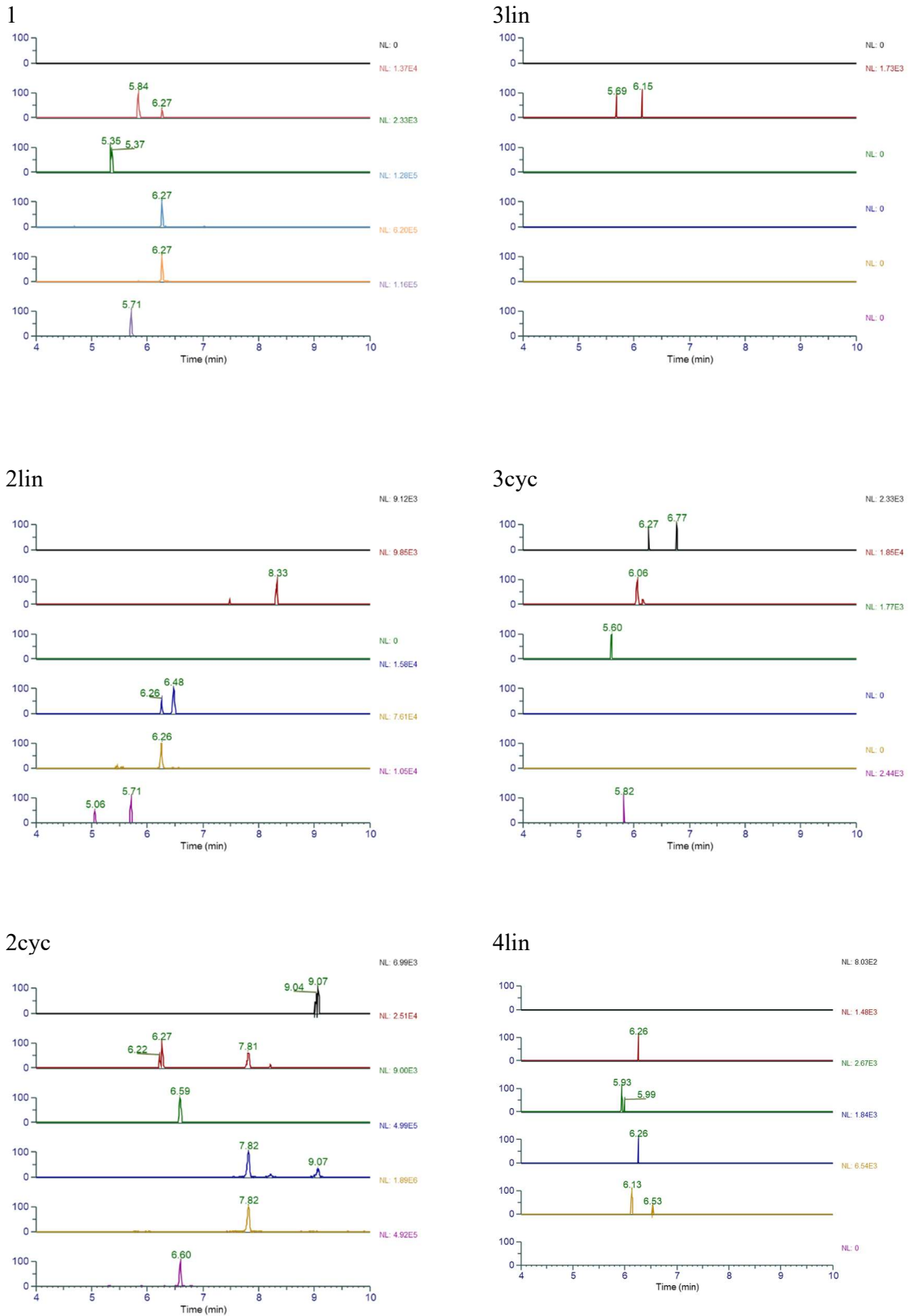
Figure S36. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-PelD9-C.

4cyc
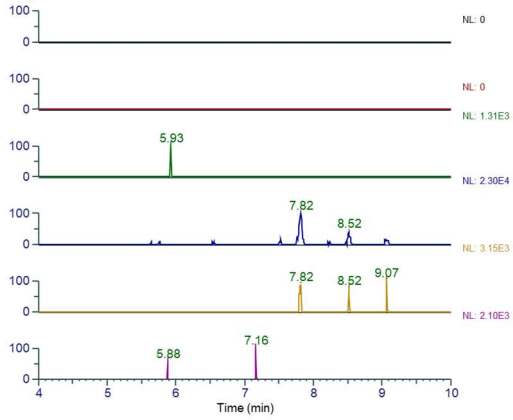
5lin
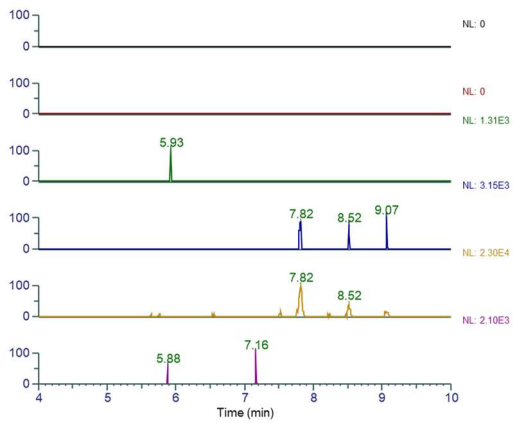
5cyc
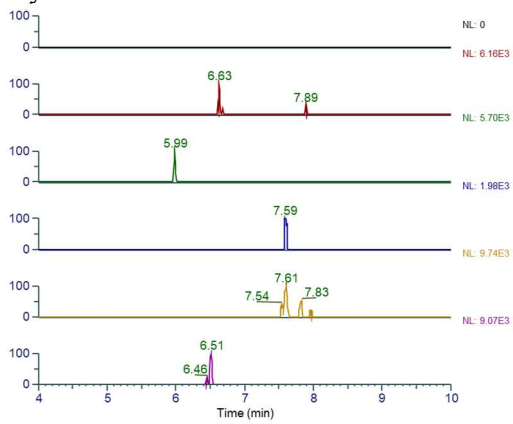
Figure S36 continued.
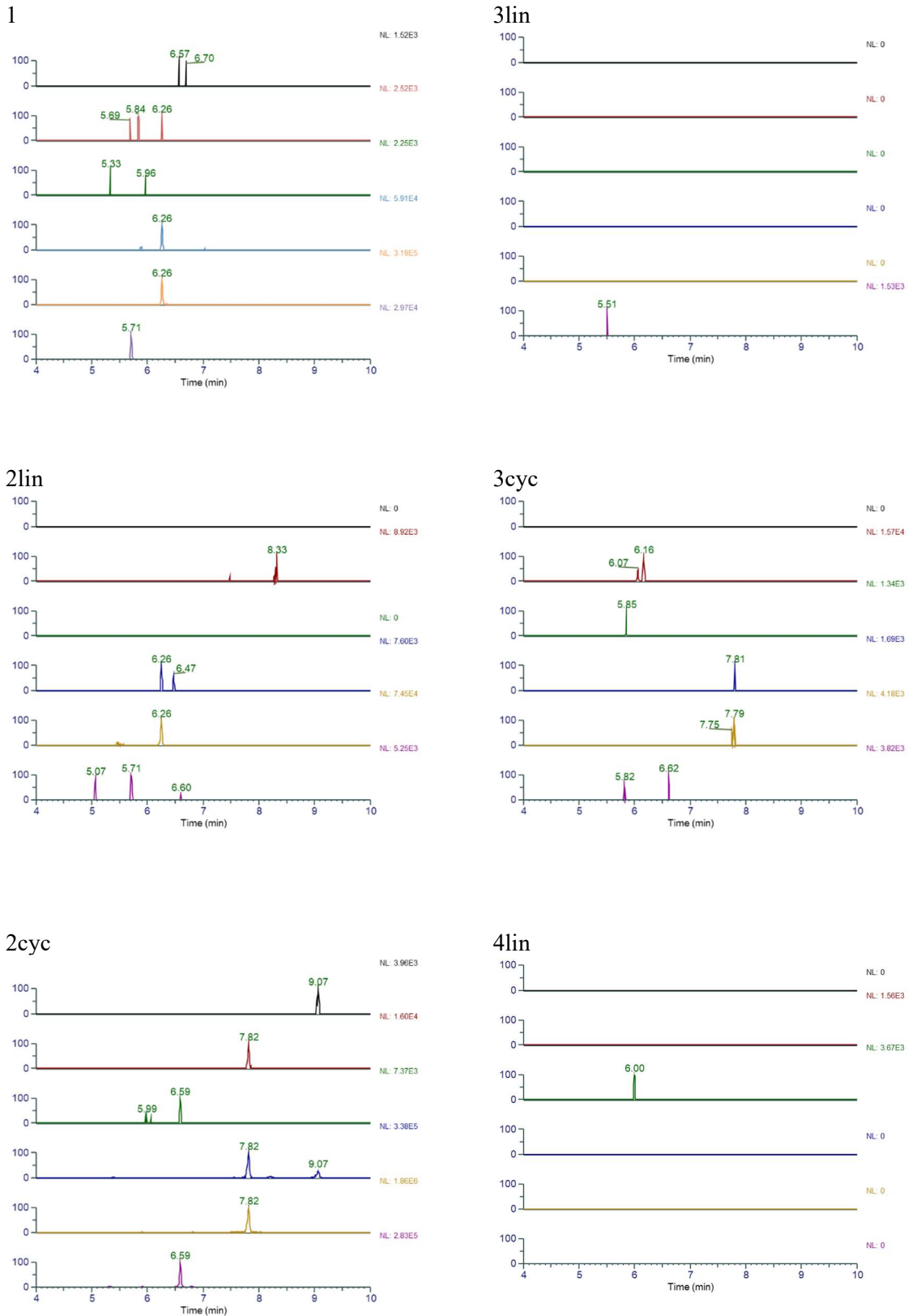
Figure S37. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-LcnB13-C.

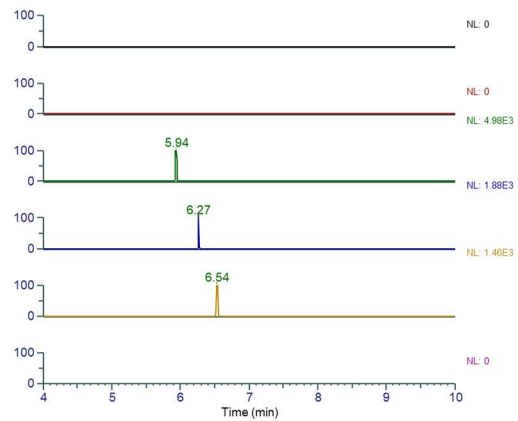4cyc
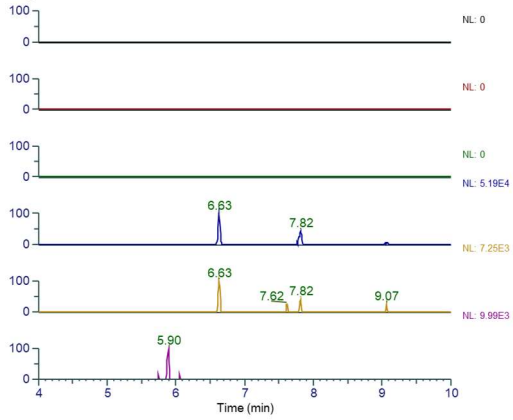
5lin
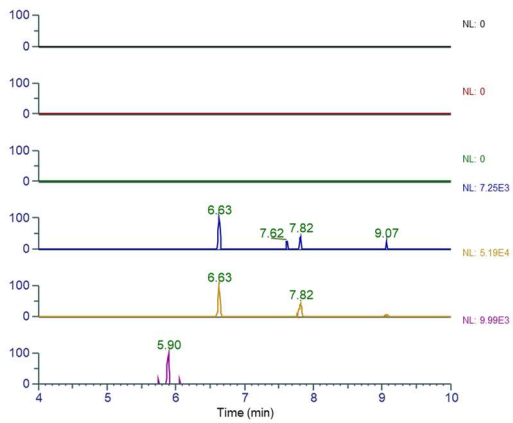
5cyc
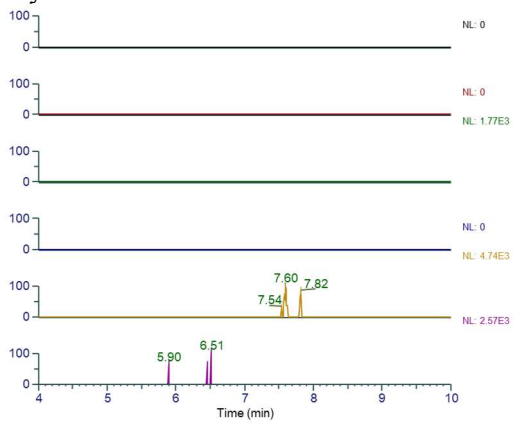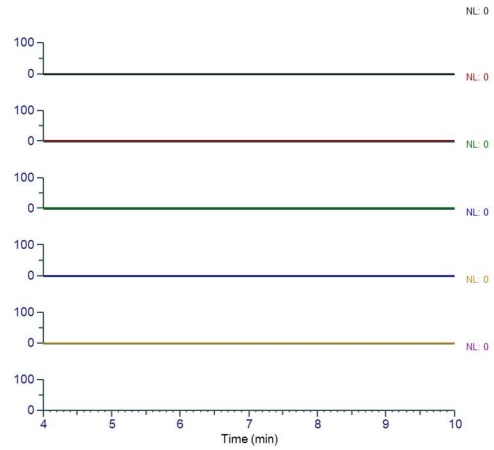
Figure S37 continued.

238
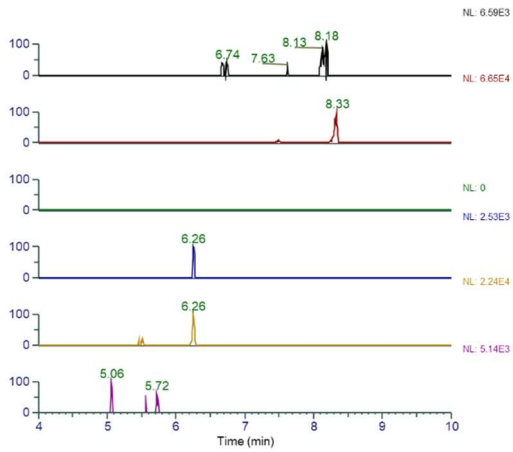
Figure S38. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-Lbm11m-C.

4cyc
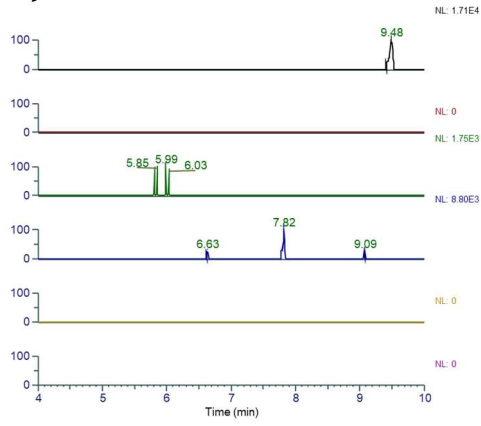
5lin
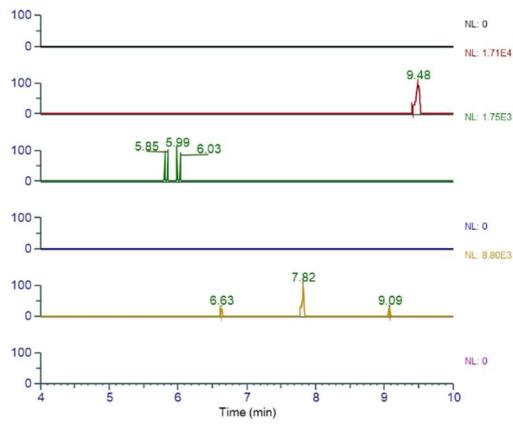
5cyc
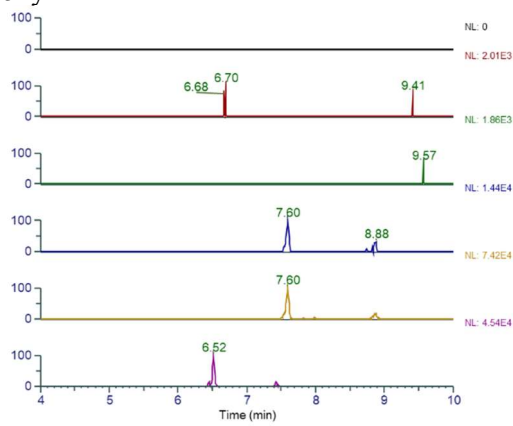
Figure S38 continued.
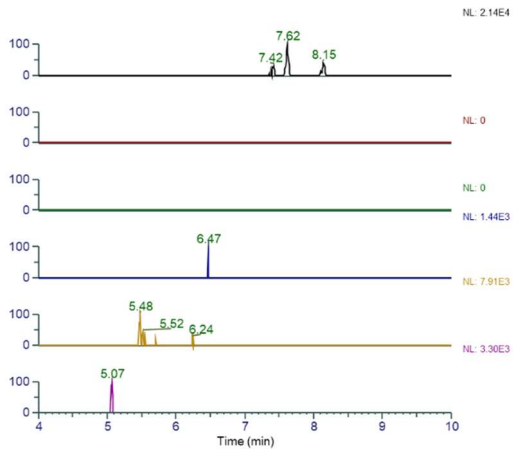
Figure S39. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-PksL5m-C.

4cyc
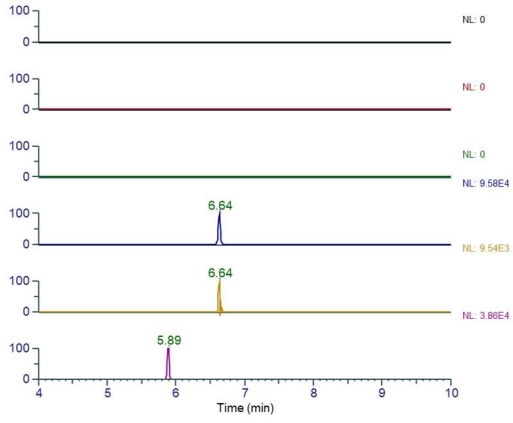
5lin
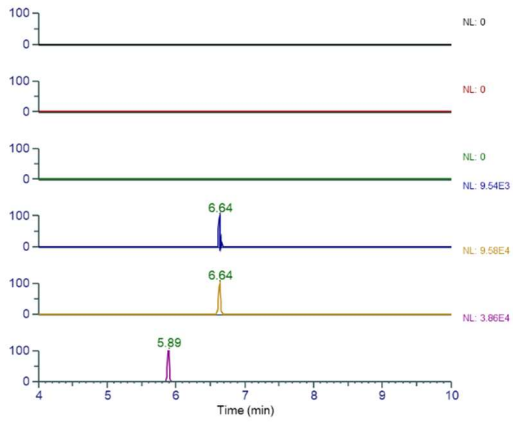
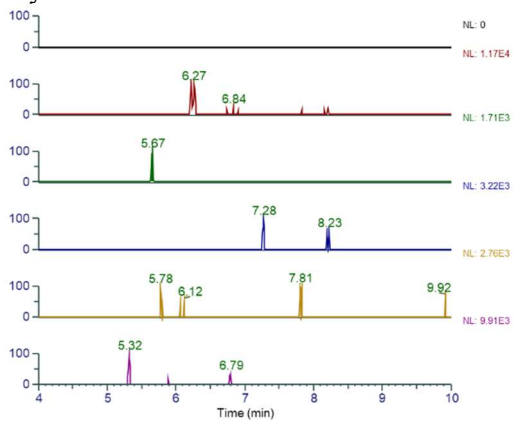5cyc
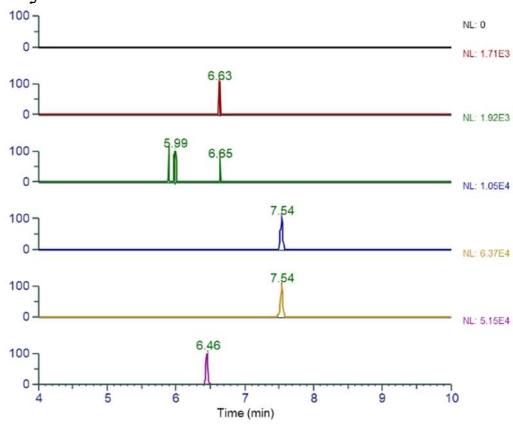
Figure S39 continued.
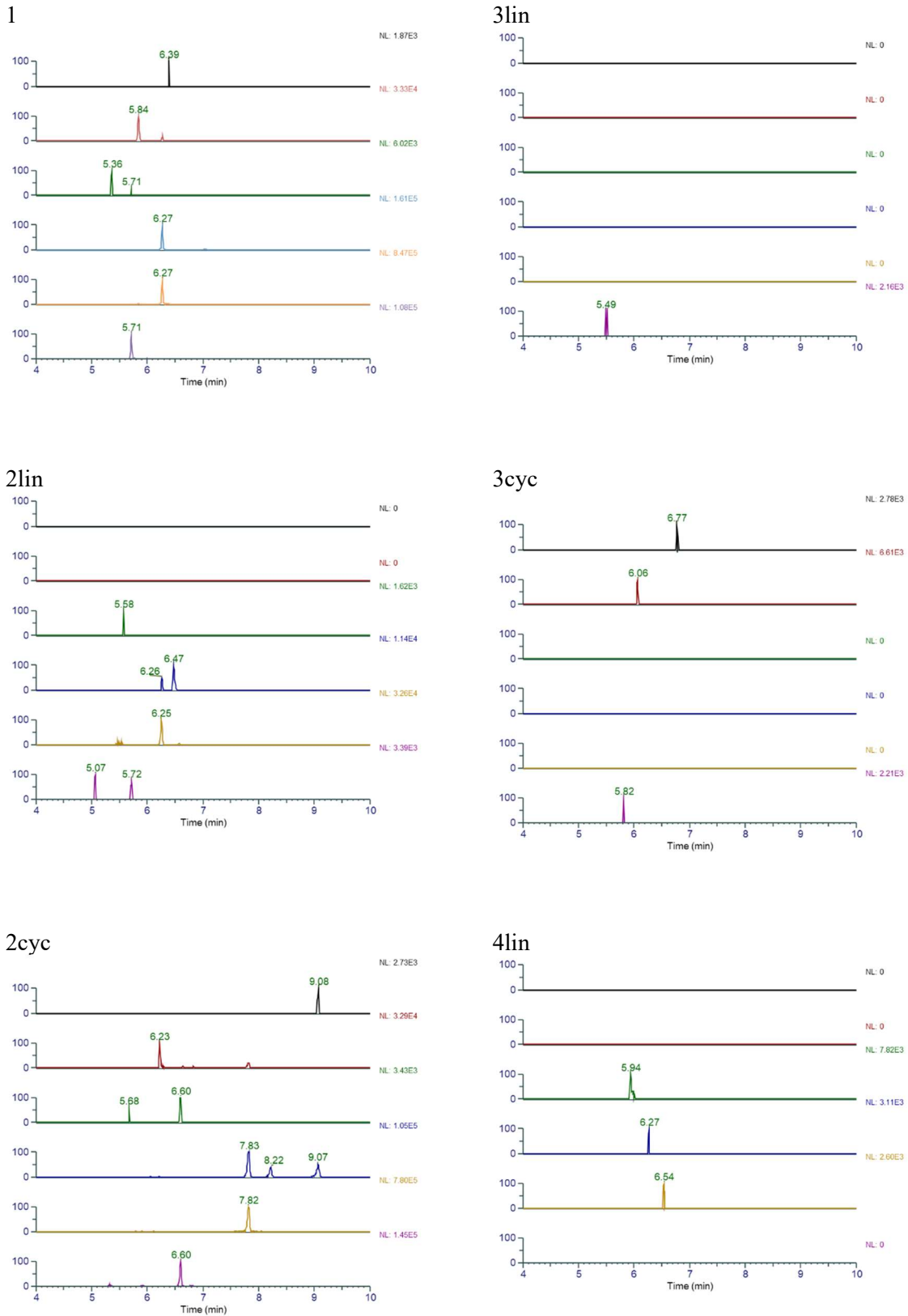
Figure S40. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-GynG11-C.

4cyc
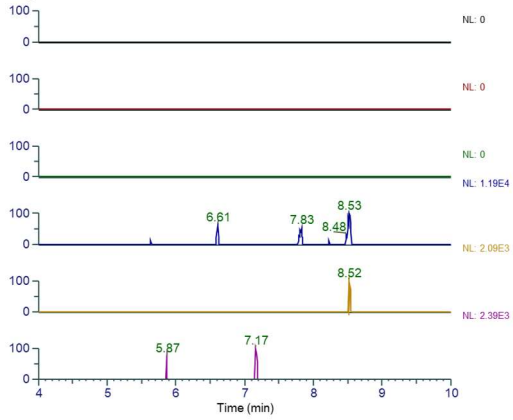
5lin
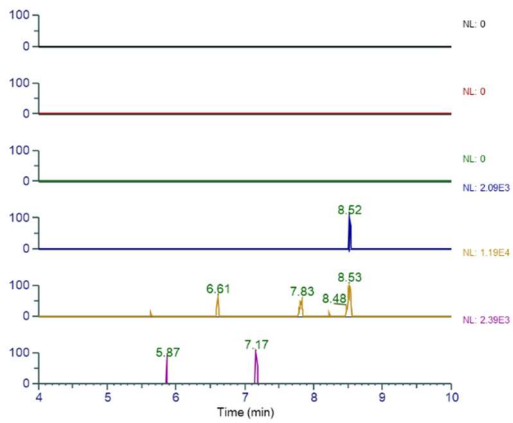
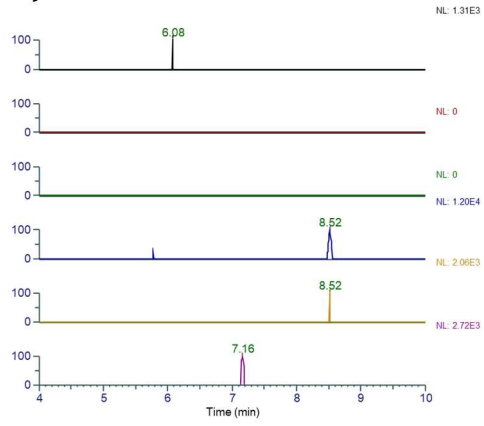5cyc
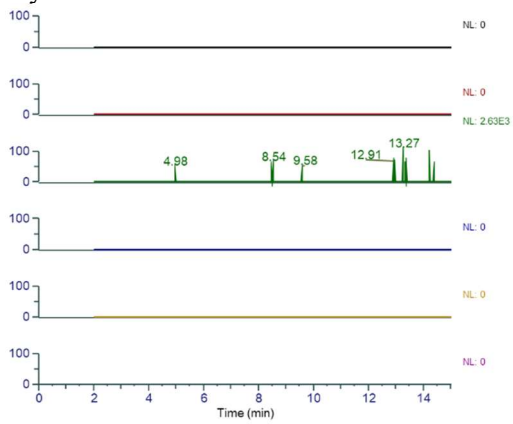
Figure S40 continued.

244
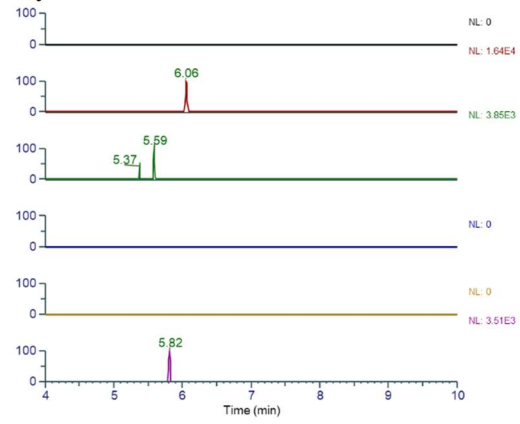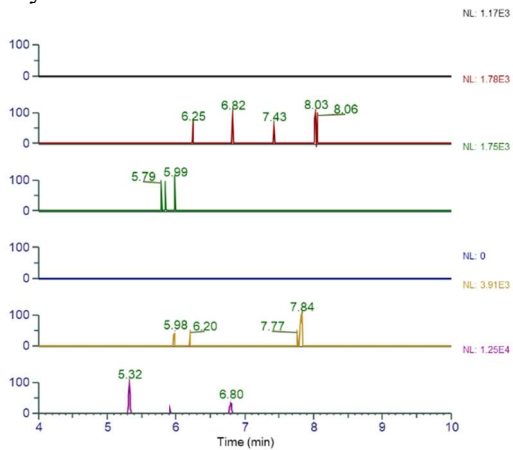
Figure S41. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-LcnA6-C.

4cyc

5lin

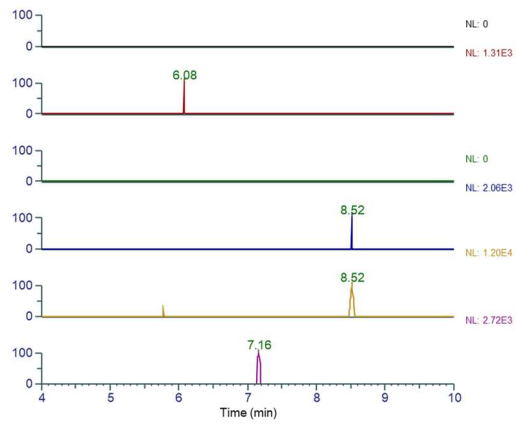

5cyc
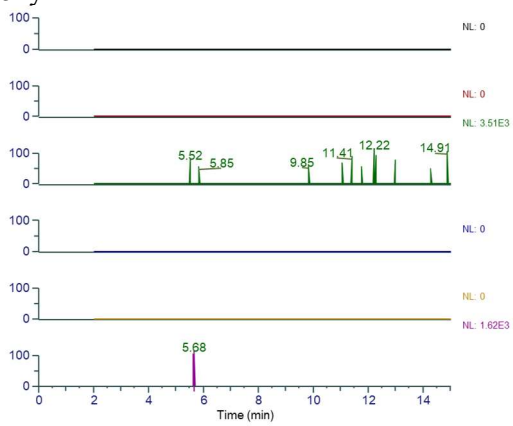
Figure S41 continued.
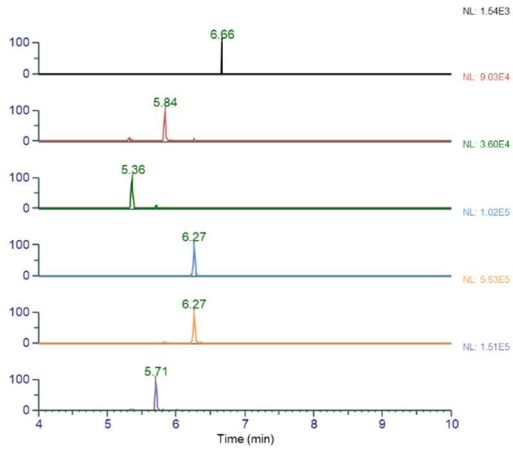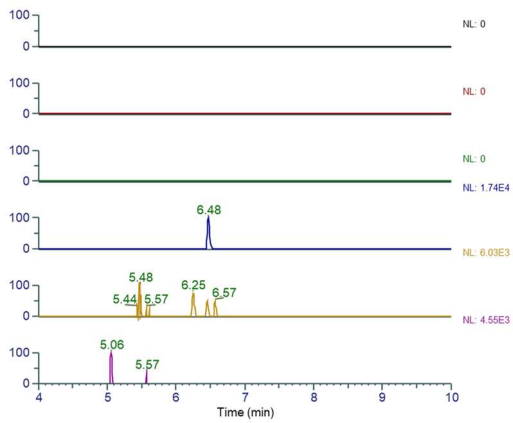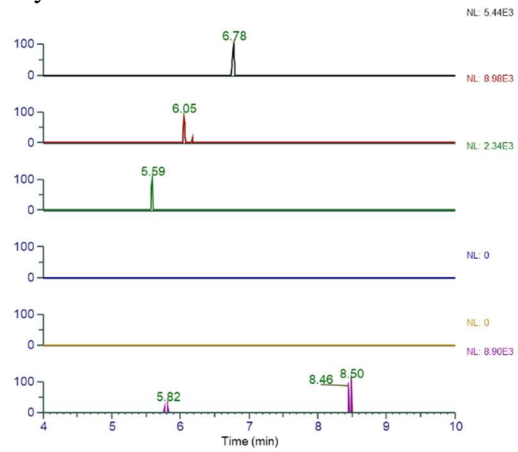
Figure S42. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-LcnB13.

4cyc

5lin

5cyc

Figure S42 continued.
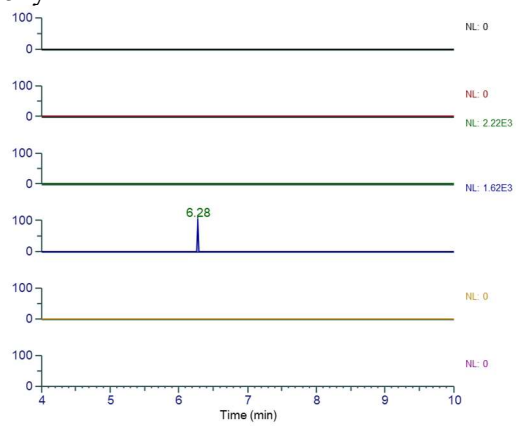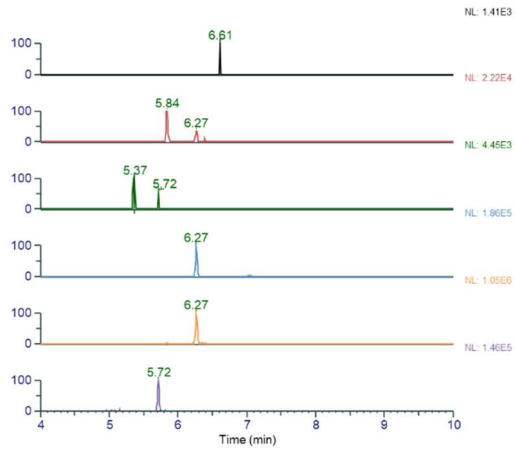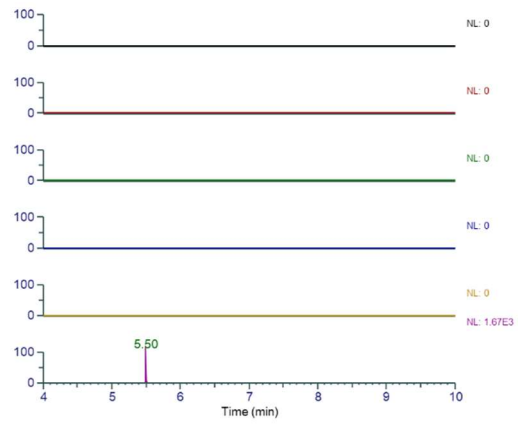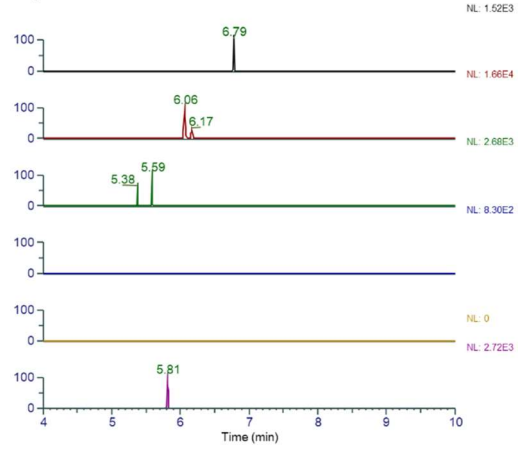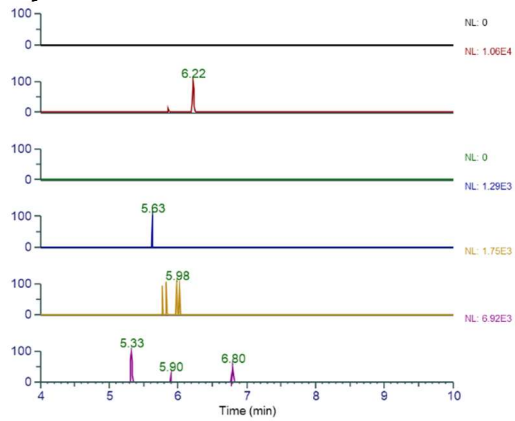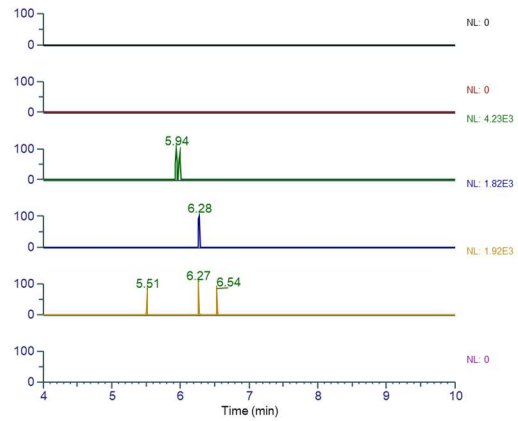
Figure S43. Extracted ion chromatograms for *S. plymuthica* 4Rx13 *ΔoocQR* harboring pBAD_OocQR-LcnB13-BasC.

4cyc
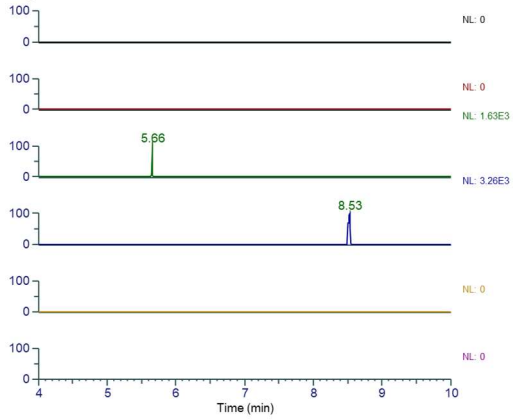
5lin
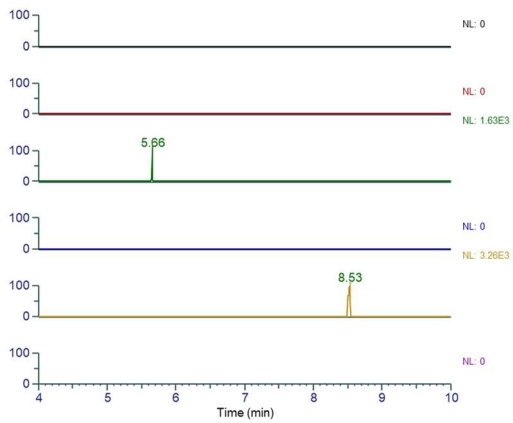
5cyc

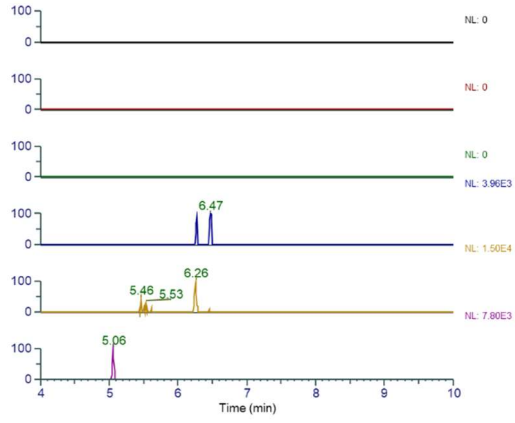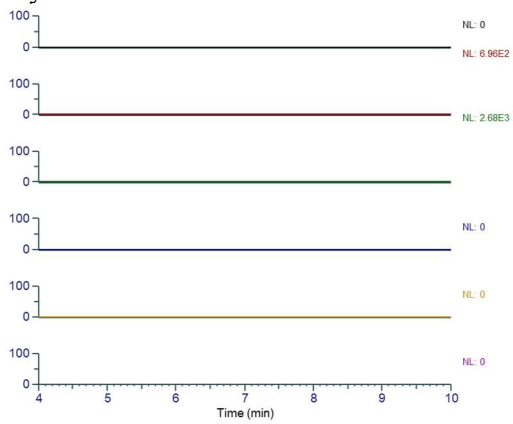**Figure S43** continued.
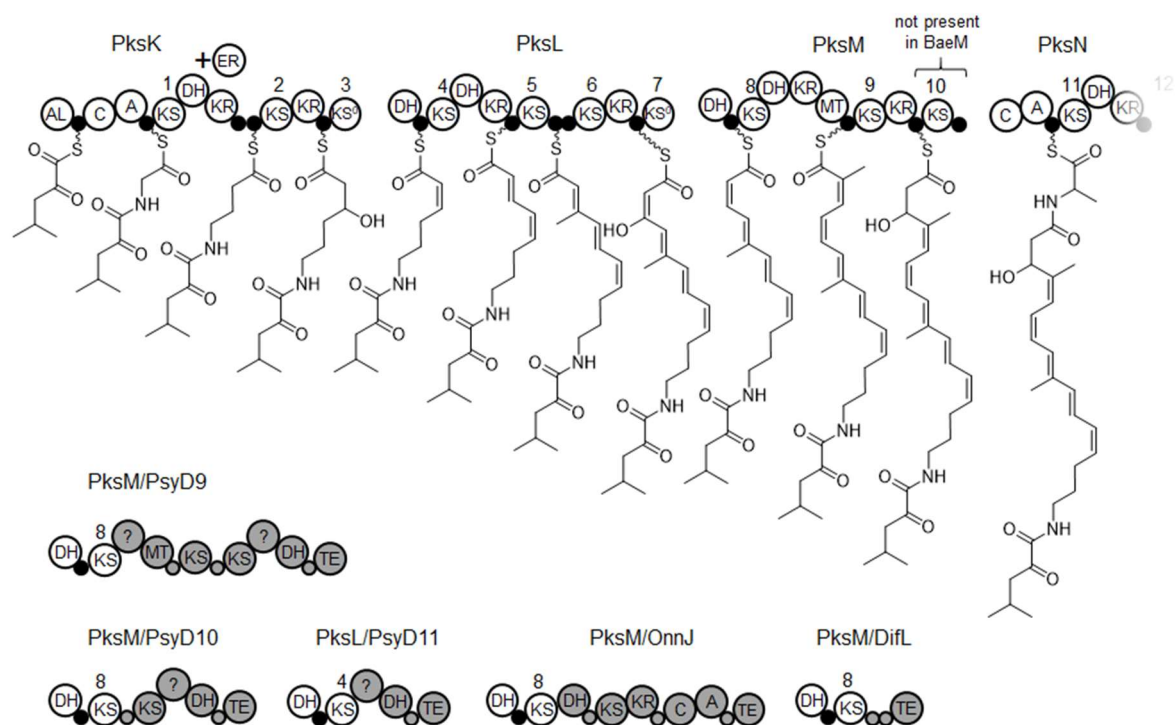
Figure S44. Overview of hybrid *trans*-AT PKSs produced by exchange of terminal modules. A) Relevant section of the bacillaene *trans*-AT PKS from *B. subtilis* (top) and hybrids created by fusion of either KS4 (located in PksL) or KS8 (PksM) with terminal modules of the psymberin (*psy*), onnamide (*onn*), or difficidin (*dif*) *trans*-AT PKSs. AL: acyl ligase, C: condensation domain, A: adenylation domain, KS: ketosynthase, DH: dehydratase, ER: enoylreductase, KR: ketoreductase, KS^0: non-elongating ketosynthase, MT: methyltransferase, filled circles: carrier protein, ?: domain of unknown function, TE: thioesterase.

References

[1]   R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792-1797.
[2]   M. N. Price, P. S. Dehal, A. P. Arkin, *PLoS One* **2010**, *5*, e9490.
[3]   G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, *Genome research* **2004**, *14*, 1188-1190.
[4]   M. A. Konkol, K. M. Blair, D. B. Kearns, *J. Bacteriol.* **2013**, *195*, 4085-4093.
[5]   D. Domik, A. Thürmer, T. Weise, W. Brandt, R. Daniel, B. Piechulla, *Front. Microbiol.* **2016**, *7*, 737.
[6]   R. Ueoka, A. Bhushan, S. I. Probst, W. M. Bray, R. S. Lokey, R. G. Linington, J. Piel, *Angew. Chem. Int. Ed.* **2018**, *57*, 14519-14523.
[7]   R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem. Int. Ed.* **2020**, *59*, 7761-7765.
[8]   E. J. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, *Nat. Commun.* **2021**, *12*, 1-14.
[9]   M. Rust, E. J. Helfrich, M. F. Freeman, P. Nanudorn, C. M. Field, C. Rückert, T. Kündig, M. J. Page, V. L. Webb, J. Kalinowski, *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 9508-9518.
[10]   a) R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straightt, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506-1509; b) J. Moldenhauer, X. H. Chen, R. Borriss, J. Piel, *Angew. Chem. Int. Ed.* **2007**, *46*, 8195-8197.
[11]   C. M. Theodore, B. W. Stamps, J. B. King, L. S. Price, D. R. Powell, B. S. Stevenson, R. H. Cichewicz, *PLoS One* **2014**, *9*, e90124.
[12]   K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews, J. Piel, *Nat. Chem. Biol.* **2009**, *5*, 494-501.
[13]   a) C. Engler, S. Marillonnet, in *cDNA libraries*, Springer, **2011**, pp. 167-181; b) T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, D. F. Voytas, *Nucleic Acids Res.* **2011**, *39*, e82-e82.
[14]   K. Blin, M. H. Medema, R. Kottmann, S. Y. Lee, T. Weber, *Nucleic Acids Res.* **2016**, gkw960.
[15]   K. Katoh, K.-i. Kuma, H. Toh, T. Miyata, *Nucleic Acids Res.* **2005**, *33*, 511-518.
[16]   F. Sievers, D. G. Higgins, in *Multiple Sequence Alignment*, Springer, **2021**, pp. 3-16.
[17]   O. Rivoire, K. A. Reynolds, R. Ranganathan, *PLoS Comput. Biol.* **2016**, *12*, e1004817.

Chapter V

Conclusions and Outlook

Conclusions and Outlook

Natural products are chemical messengers involved in the communication between organisms.[1] Numerous different interactions have been described, ranging from the coordination of processes among members of the same species to inter-kingdom cross-talk in symbiosis and pathogenicity.[2] The functions of natural products in these interactions are broad and include color or fragrance changes to either repel enemies or attract pollinators as well as compounds that fight off attacking organisms.[3] These highly diverse activities are fascinating and have ignited studies towards human applications. Especially their function in the context of attack or defense has been employed by humans in the form of antibiotics. While the first antibiotics a few millennia ago were mainly based on plant extracts, Fleming's discovery of penicillin and the subsequent large scale production in the early 20th century spurred an entire era of drug discovery.[4] During the "Golden Age of Antibiotics" readily-accessible fungal and bacterial strains were extensively studied and led to the isolation of numerous bioactive molecules. But after an initial rise in the number of discoveries, rediscoveries became more frequent and the number of newly found antibiotic scaffolds stagnated.[5] As such, this may not seem problematic, but the misuse of antibiotics over the last decades and the continuous exposure of microorganisms toward these drugs, has led to the evolution of resistant strains. The combination of multi-drug resistant bacteria and the decreasing number of new antibiotics has led to a situation where we cannot necessarily rely on antibiotics for the treatment of microbial infections anymore.

One way to tackle these problems is through synthetic organic chemistry. In combinatorial chemistry, large libraries of synthetic compounds are generated and tested for bioactivity. The hit rate to find bioactive molecules in these libraries is, however, surprisingly low.[6] Studies on the chemical space occupied by bioactive natural products and randomly produced synthetic compounds has revealed differences that could explain the inefficiency of such approaches to obtain novel antibiotics.[7] One way to enhance the hit rate among synthetic compounds therefore relies on the derivatization of natural products. This strategy is commonly employed nowadays and most drugs on the market are either natural products or natural product-derived. In diversity-oriented synthesis natural product scaffolds are used as a starting point.[8] The target compound is diversified to improve its stability or membrane permeability along with other important pharmacological properties. Sometimes simple modifications are sufficient – acetylation of salicylic acid provides us with aspirin, for example. [9] More commonly though, modifications involve complex synthetic procedures. Further, natural products are often not isolated in large quantities from their natural source which may itself also be hard to access. Even though advanced chemical syntheses have been successfully used for total syntheses or to produce natural product-derived drugs, they are often laborious and difficult to design.[8] A closer look at the biosynthesis of natural products can provide insights into how nature produces these highly specialized metabolites.[10]

One major class of antibiotics are polyketides. Their underlying biosynthetic machinery resembles an assembly line in a factory where an intermediate is channeled through a series of modules. Each module introduces a set of modifications and brings the intermediate closer to the final product. In the case of polyketide biosynthesis, the assembly lines are called polyketide synthases (PKSs).[11] In general, the order and composition of the modules correlates with the structure of the final compound. This relationship can be used to predict a polyketide product from a PKS, or *vice versa*, to infer a PKS module sequence from the polyketide and is referred to as the collinearity rule.[12] In the subgroup of *trans*-acyltransferase PKSs (*trans*-AT PKSs), the biosynthesis is getting more complex as the assembly line can be expanded by stand-alone domains that act *in trans*. However, phylogenetic analyses of the ketosynthases (KSs) have revealed a remarkable feature in that they clade with other KSs that accept intermediates with the same chemical properties in the α-β position.[13] This attribute was termed the *trans*-AT PKS correlation rule and can be used to predict final polyketide products from *trans*-AT PKSs.

This was implemented into the bioinformatics tool transATor for automated, KS-based compound prediction.[14]

For precise prediction of *trans*-AT PKS-derived polyketides, the transATor algorithm depends on a collection of KSs from characterized biosynthetic models. Based on these models, the specific modifications present in the intermediate a KS accepts are defined. These assigned "specificities" in turn define the phylogenetic clades. When confronted with a KS query, the modifications occurring in the module upstream of the KS are predicted based on the name of the phylogenetic clade the query falls into.[14] To improve the predictability and to reduce the number of undefined clades, a large enough collection of annotated *trans*-AT PKS biosynthetic models is necessary. Along with KSs that follow canonical, frequently observed PKS modules, the exploration of KSs downstream of unusual module architectures is especially important. If not present in the training set of the prediction tool, the query of such KSs would lead to orphan clades or ambiguous assignments. In Chapter II, we characterized the unusual starting modules of the *trans*-AT PKS for lobatamide. In earlier studies, the biosynthetic gene cluster was assigned to the final compound, but the transformations occurring in the starting modules remained elusive.[15] Based on the final compound, these should be responsible for the installation of a methylated oxime moiety, a functional group not commonly found in polyketides. As expected for unusual chemical modifications, the respective KSs fell into orphan clades. We therefore tested the oxygenase encoded in the starting modules of the Lbm PKS on substrate mimics and were able to confirm its role in oxime formation. The oxygenase was able to install oxime groups on non-native substrates, making it a potential tool for biocatalytic purposes. We further heterologously expressed the starting modules *in vivo* and successfully produced the methoxime intermediate expected based on the biosynthetic model. The functional expression of this series of *trans*-AT PKS modules is not only a potential starting module for engineered oxime-harboring polyketides, but also confirms the biosynthetic model for lobatamide biosynthesis. Lastly, the characterization of these oxime-forming and methylating modules and the assignment of the two KSs as oxime- and methoxime-accepting, respectively, increase the described biocatalytic potential of *trans*-AT PKSs and simultaneously improve the predictive power of tools like transATor.

While *trans*-AT PKS-derived domains may find applications in the context of synthetic organic chemistry, the multimodular organization of PKSs themselves makes them an interesting target for engineering. In theory, it should be possible to reorganize the modules within an assembly line to produce new, designed polyketide products. With this in mind, extensive studies have been performed in the *trans*-AT PKS-related *cis*-AT PKSs, but mainly with low yields.[16] In addition to architectural factors, the evolution behind *cis*-AT and *trans*-AT PKSs exhibit major differences. Compared to vertical gene transfer and gene duplication in *cis*-AT PKSs, extensive horizontal gene transfer is observed for *trans*-AT PKSs.[13] Such an extensive horizontal gene transfer results in mosaic-like *trans*-AT PKSs that harbor module units from various sources.[13] This natural reshuffling of *trans*-AT PKS parts is in line with the correlation rule, which suggests that KSs co-evolve with upstream domains. The occurrence of exchanged module blocks raises the question of whether these blocks are independent of the original PKS and functional when integrated into another *trans*-AT PKS. By only accepting intermediates that have undergone modifications according to the preceding domains, the KS may act as a gatekeeper to ensure the correct assembly of the final polyketide. The independent and mobile nature of these *trans*-AT PKS module blocks would make them predestined for engineering efforts. Interestingly, even the exchange of entire module series across *trans*-AT PKSs can be found in nature.[17] This further substantiates their role as candidates for artificial reprogramming purposes.

We embarked on our first *trans*-AT PKS engineering adventures in Chapter III. To identify an initial set of basic principles for *trans*-AT PKS recombineering, we focused on their offloading domains. The functionality of these releasing domains can be rather easily assessed in the presence of absence of the

final polyketide. Using the well-characterized biosynthetic assembly line for bacillaene in *Bacillus subtilis*, we exchanged the terminal domains with the corresponding domains from other PKSs. The offloading efficacy was measured within the native host based on bacillaene yield. As important factors affecting the release of the final intermediate from the bacillaene *trans*-AT PKS, we identified the releasing domain's mode of action as well as the chemical properties of the intermediate. In addition, a *trans*-acting condensation domain was found to increase release of bacillaene. These results underpin the fact that the compatibility of chemical properties between native and non-native polyketide intermediates is pivotal for the generation of functional hybrid assembly lines.

Chapter IV continued to explore engineering of *trans*-AT PKSs in the highly versatile biosynthetic pathway for oocydins in *S. plymuthica* 4Rx13. The Ooc *trans*-AT PKS may be the most complex PKS described to date and harbors an unusual halogenation module.[18] Pull-down assays showed that within this module, a halogenase associates with an auxiliary protein to chlorinate the growing polyketide chain *in trans*. In a flexible set up, we exploited this feature and introduced non-native PKS modules on a plasmid. As an engineering platform, we constructed a *Serratia* mutant deficient in the auxiliary protein and the PKS modules downstream thereof but still contained the halogenase gene encoded directly upstream of the auxiliary protein gene. The mutant strain was complemented with a plasmid harboring the auxiliary protein and additional downstream modules. Upon cultivation of these strains and expression of the genes for the introduced plasmid-borne proteins, the newly introduced PKS parts interacted closely enough through the halogenase-auxiliary protein interaction to allow processing of the substrate. Using this setup, we tested a range of hybrid constructs and assessed their ability to produce new successfully elongated and chlorinated, truncated oocydin congeners. Conveniently, the characteristic isotope pattern upon chlorination allowed the precise identification of the formed intermediates by mass spectrometry, even at low abundance. One main question regarding the engineering of hybrid PKSs was whether and where native and non-native PKS parts can be recombined. The evolutionary studies mentioned earlier suggested rewiring of a region downstream of the KS as a promising strategy. To obtain deeper insights into the co-evolution of the domains around this region, we employed a statistical approach. Statistical coupling analysis of extracted domain sequences around the KSs from all *trans*-AT PKS clusters deposited in the antiSMASH database identified a possible fusion site at the C-terminus of a flanking subdomain region following the KS. The fusion site was validated in experiments that introduced non-native modules. Additionally, a fusion site directly C-terminal to the KS but upstream of the flanking subdomain region was tested and found to be inactive. These results implied that the flanking subdomain region following the KS is important for the proper functioning of the respective module. Interestingly, the conserved motif used as a fusion site is also present in *cis*-AT PKSs, which harbor an AT instead of a flanking subdomain in this region.[19] Even though structural information on *trans*-AT PKSs remains sparse, structural modelling of the KS and the following flanking subdomain suggested that the subdomain forms a defined structure where the AT would be found in *cis*-AT PKSs. This region may thus be important for protein-protein interactions between the domains within a module. An impaired or non-native flanking subdomain may impede the shuttling of the ACP-bound intermediate to the different active sites.[20] Furthermore, the conserved motif downstream of the flanking subdomain was modelled to lie on the surface of the KS facing towards the first domain of the following module. The end of the motif thus represents an accessible region that connects the flanking subdomain of one module with the following module. These findings uncovered a way to generate engineered *trans*-AT PKSs in a strategy guided by evolutionary principles. The lack of structural information on these megaenzyme complexes can at least be partially overcome by deep statistical analysis, phylogenetic analyses, and logical reasoning.

In conclusion, the goal of this thesis was to identify basic principles for the engineering of bacterial hybrid *trans*-AT PKSs. Polyketides represent a major part of drugs but are often highly complex. Despite advances in synthetic organic chemistry, synthetic routes toward these elaborate compounds are often

laborious and complicated. Considering the large enzymatic toolbox occurring naturally in *trans*-AT PKSs and their modular organization, these biofactories are predestined for reprogramming. Such artificially rewired bacterial assembly lines would produce designed polyketides in a fast and economical way. Even though the design of hybrid PKSs to produce target polyketides appears simple on paper, many factors influence the processivity of these enzymatic machineries. To obtain easily recombinable module blocks for artificial assembly lines with high enough throughput, further insights need to be gathered. Recent findings from evolutionary, biochemical, and structural biology are gradually uncovering the mysteries surrounding these megaenzymes. Even if it still seems like a futuristic endeavor, we are inching closer toward the ultimate goal of bacterial production of pharmaceutically relevant polyketides through reprogrammed *trans*-AT PKSs.

References

[1]   D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2020**, *83*, 770-803.

[2]   a) F. Pankewitz, M. Hilker, *Biol. Rev.* **2008**, *83*, 209-226; b) R. Bandichhor, B. Nosse, O. Reiser, *Nat. Prod.Synth.* **2004**, 43-72.

[3]   a) M. Horikawa, M. Shimazu, M. Aibe, H. Kaku, M. Inai, T. Tsunoda, *J. Antibiot. Res.* **2018**, *71*, 992-999; b) N. Nikoh, T. Tsuchida, T. Maeda, K. Yamaguchi, S. Shigenobu, R. Koga, T. Fukatsu, E. G. Ruby, G. Bennett, Y. Hongoh, *mBio* **2018**, *9*, e00890-00818.

[4]   D. A. Dias, S. Urban, U. Roessner, *Metabolites* **2012**, *2*, 303-336.

[5]   G. M. Cragg, D. J. Newman, *Pure Appl. Chem.* **2005**, *77*, 7-24.

[6]   I. Paterson, E. A. Anderson, *Science* **2005**, *310*, 451-453.

[7]   T. H. Keller, A. Pichota, Z. Yin, *Curr. Opin. Chem. Biol.* **2006**, *10*, 357-361.

[8]   W. R. J. D. Galloway, A. Isidro-Llobet, D. R. Spring, *Nat. Commun.* **2010**, *1*, 80.

[9]   R. J. Spandl, A. Bender, D. R. Spring, *Org. Biomol. Chem.* **2008**, *6*, 1149-1158.

[10]   T. A. Scott, J. Piel, *Nat. Rev. Chem.* **2019**, *3*, 404-425.

[11]   E. J. N. Helfrich, J. Piel, *Nat. Prod. Rep.* **2016**, *33*, 231-316.

[12]   C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688-4716.

[13]   T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225-233.

[14]   E. J. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, *Nat. Chem. Biol.* **2019**, *15*, 813-821.

[15]   R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem. Int. Ed.* **2020**, *59*, 7761-7765.

[16]   A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, *Chem. Rev.* **2019**, *119*, 12524-12547.

[17]   E. J. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, *Nat. Commun.* **2021**, *12*, 1-14.

[18]   F. Hemmerling, R. A. Meoded, A. E. Fraley, H. A. Minas, C. L. Dieterich, M. Rust, R. Ueoka, K. Jensen, E. J. Helfrich, C. Bergande, M. Biedermann, N. Magnus, B. Piechulla, J. Piel, *Angew. Chem. Int. Ed.* **2022**.

[19]   J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380-416.

[20]   a) K. J. Weissman, *Nat. Chem. Biol.* **2015**, *11*, 660-670; b) D. P. Cogan, K. Zhang, X. Li, S. Li, G. D. Pintilie, S.-H. Roh, C. S. Craik, W. Chiu, C. Khosla, *Science* **2021**, *374*, 729-734.